

User-centred redesign of a speech research database web
interface and its associated app

A thesis
submitted in partial fulfilment
of the requirements for the Degree
of
Masters of Human Interface Technology
by
Qi Min Ser

Supervisors:
Senior Supervisor: Dr. Christoph Bartneck¹
Co-supervisor: Dr. Gun Lee¹

University of Canterbury
August 2016

¹ Human Interface Technology Laboratory New Zealand, University of Canterbury

Abstract

This thesis project aim to use a user-centred design approach to redesign two related applications, LaBB-CAT and ElicitSpeech. LaBB-CAT is a speech research database that allows researchers to easily search through the trove of corpus data, listen to speech recordings, and export demographic information of the speaker as well as transcript information and metadata. Through the requirements gathering process, it was found that the information architecture was confusing and difficult to understand. A new search methodology was therefore implemented, and several features of the information architecture were improved. Although the redesign process appeared to have improved the usability of the interface, aspects that can be further improved are mentioned.

ElicitSpeech is a speech recording application that elicits speech from users to be recorded for research purposes. The requirements gathered for this system found that the current system is easy to learn and use, but the motivation and attractiveness of the system requires improvement. Gamification elements were added and similar applications were consulted for the design of the final prototype. A statistically significant improvement in attractiveness and efficiency was found after the redesign.

This thesis project was able to create a usable interface of a speech research database, as well as a user-friendly interface of a speech recording application, which would contribute to hastening the progress of future speech and linguistics research. Future work for the redesign project of LaBB-CAT should focus on the learnability of the interface as well as the upload feature of the interface. Future redesign projects for ElicitSpeech should focus on designing for the motoric disabilities that people with dysarthria might have, as well as to incorporate gamification elements to improve user motivation.

Table of Contents

Abstract	i
Acknowledgments	v
Chapter 1: Introduction	1
1.1 Motivation	2
1.2 Objectives	3
1.3 Scope	3
1.4 Structure of Thesis	4
Chapter 2: Theory	5
2.1 User-Centred Design	5
2.2 User Experience	6
2.3 Information architecture	7
2.4 Gamification	10
2.5 Summary	13
Chapter 3: LaBB-CAT Requirements Gathering	14
3.1 Background of LaBB-CAT	14
3.2 Current web interface of LaBB-CAT	15
3.3 Usability testing of LaBB-CAT's current interface	17
3.3.1 Participants	17
3.3.2 Design and Apparatus	18
3.3.3 Procedure	18
3.4 Results	19
3.5 Summary	20
Chapter 4: LaBB-CAT Redesign Phase 1	21
4.1 Design	21
4.2 Low-fidelity prototyping of LaBB-CAT	24

4.3	Usability testing of LaBB-CAT's low-fidelity prototype	28
4.3.1	Participants	28
4.3.2	Design and Apparatus	28
4.3.3	Procedure	29
4.4	Results	30
4.5	Summary	31
Chapter 5:	LaBB-CAT Redesign Phase 2	33
5.1	Design	33
5.2	High-fidelity prototyping of LaBB-CAT	34
5.3	Usability testing of LaBB-CAT's high-fidelity prototype	34
5.3.1	Participants	35
5.3.2	Design and Apparatus	36
5.3.3	Procedure	36
5.4	Results	37
5.5	Summary	39
Chapter 6:	ElicitSpeech Requirements Gathering	40
6.1	Background of ElicitSpeech	40
6.2	Current web interface of ElicitSpeech	41
6.3	Usability testing of ElicitSpeech's current interface	41
6.3.1	Participants	42
6.3.2	Design and Apparatus	44
6.3.3	Procedure	44
6.4	Results	44
6.5	Summary	45
Chapter 7:	ElicitSpeech Redesign Phase 1	47
7.1	Design	47
7.2	Low-fidelity prototyping of ElicitSpeech	48
7.3	Usability testing of ElicitSpeech's low-fidelity prototype	49
7.3.1	Participants	49
7.3.2	Design and Apparatus	50
7.3.3	Procedure	50
7.4	Results	52

7.5	Summary	53
Chapter 8:	ElicitSpeech Redesign Phase 2	54
8.1	Design	54
8.2	Second low-fidelity prototyping of ElicitSpeech	57
8.3	Usability testing of ElicitSpeech's second low-fidelity prototype	57
8.3.1	Participants	57
8.3.2	Design and Apparatus	59
8.3.3	Procedure	61
8.4	Results	62
8.5	Summary	63
Chapter 9:	ElicitSpeech Redesign Phase 3	64
9.1	Design	64
9.2	High-fidelity prototyping of ElicitSpeech	65
9.3	Experimental evaluation of ElicitSpeech's high-fidelity prototype	65
9.3.1	Setup	65
9.3.2	Participants	68
9.4	Results and Discussion	68
9.4.1	Comparison of data obtained from CrowdFlower and FaceBook	68
9.4.2	Comparison of Old interface and New prototype	71
9.4.3	Benchmark	77
9.5	Summary	78
Chapter 10:	Conclusion	80
10.1	Summary	80
10.2	User-centred design methodology	81
10.3	Outlook	82
10.3.1	LaBB-CAT	82
10.3.2	ElicitSpeech	85
10.4	Visions for the future	87
	References	88

Appendix A: Appendix A: Interview script used in user study for LaBB-CAT	92
Appendix B: Appendix B: Questionnaires used in user study for LaBB-CAT	95
Appendix C: Appendix C: Interview script in user study for ElicitSpeech	97
Appendix D: Appendix D: UEQ used in user study 3 and 4 for ElicitSpeech	99
Appendix E: Appendix E: Information sheet and consent forms	101
Appendix F: Appendix F: Statistical tables	103

Acknowledgments

I would like to express my sincere gratitude to my senior supervisor Dr. Christoph Bartneck for his expertise and guidance in every step of my thesis project. I also wish to thank my co-supervisor Dr Gun Lee for all his encouragements, support, and insightful comments for my thesis.

I would like to thank Ken Beckman for his valuable assistance in obtaining my scholarship and for various assistance given during the study. I also thank The New Zealand Institute of Language, Brain and Behaviour (NZILBB) for providing me with my scholarship for the study. I would also like to thank Megan McAuliffe, Robert Fromont, and Jakub Zlotowski for all their support and counsel in my thesis project.

Further, I would like to thank my family in Singapore, as well as my host family in New Zealand, Jeremy and Esther, for all their patience, encouragement, and support throughout my studies. Finally, I would especially like to thank all my friends and staff at the HITLabNZ for the wonderful experiences, encouragements, and help I have received while studying there.

Chapter I

Introduction

Since the origins of human society, we have been dependent on our ability to communicate. Science and technology have sought to improve communication amongst humans, yet communication disorders can have considerable detrimental social, educational, workplace and economic effects, particularly if the conduct of day-to-day tasks is challenging.

According to the available statistics [11], one in six Americans will have a communication disorder at some point in their lives. As of 2010, there are approximately 7.5 million people in the US who have trouble with using their voice [10]. Around 9 percent of young children in the US have speech sound disorders, and approximately 5 percent of children develop noticeable speech disorders by the first grade [10]. There is still no known cause for the development of many of these disorders.

One type of communication disorder is dysarthria. Speech is usually produced by precise and coordinated muscle movements in the abdomen, chest, neck, and head [8]. Dysarthria occurs when a neurological disorder causes these muscle movements to become slow, weak and uncoordinated [8]. This results in the reduction of the intelligibility of speech and ability to communicate with others. Although some people are born with Dysarthria, which could be a side-effect of cerebral palsy, it can also be acquired later in life from diseases such as Parkinsons disease, brain injury, and stroke [8].

As dysarthria is traditionally viewed from an English-centric perspective, yet most of the world is bilingual, we need to know how dysarthria presents in different languages and how listeners of different language backgrounds process dysarthric speech. To do this, a version of LaBB-CAT (Language, Brain and Behaviour - Corpus Analysis Tool) was set up to be a cross-linguistic database for researchers to work on people with dysarthrias audio

recordings and transcripts [20, 19, 6].

The current project investigates the usability of the LaBB-CAT interface for both novice and expert researchers, and the usability of the associated app, ElicitSpeech, for non-expert users using a user-centred design methodology. ElicitSpeech would be used by the speakers who were recruited for the New Zealand Institute of Language, Brain and Behaviour (NZILBB) Dysarthria research to carry out the voice recording tasks. This can either be a mobile or web application. LaBB-CAT is essentially a web database for organisation, storage, and analysis of audio recordings received from the speakers who had submitted their voice recordings through the ElicitSpeech app.

First, requirements for the interface was gathered from the client, developer, as well as target users. After the requirements were gathered, the project was put through phases. Each phase consisted of one round of design, prototyping, and usability testing.

This introductory chapter contains four sections. The motivation and goals of this PhD research are explained in section 1.1 and 1.2, respectively. The scope of the thesis project is stated in section 1.3. Finally, section 1.4 gives an overview of the remaining chapters of this proposal.

1.1 Motivation

In the initial meetings with the business clients of the project, it was apparent that there were gaping problems in the LaBB-CAT interface. The developer mentioned that many users of LaBB-CAT send him emails asking for instructions on how to use the website. The business clients also mentioned that they would like ElicitSpeech, an application built to collect speech data for LaBB-CAT, to be as user-friendly as possible before opening them to the public. This is to ensure that users would not face problems after launching the product, which would create high costs for the business clients.

The current LaBB-CAT and ElicitSpeech interfaces have never been designed using a user-centred design approach. They would first undergo a requirements gathering process to find out the specific problems of each interface from the target users. Each interface would then be redesigned, pro-

totyped, and evaluated to ensure that the provided solution is sufficiently good for users.

1.2 Objectives

The objective of this project is to redesign the current LaBB-CAT and ElicitSpeech interfaces using a user-centred approach so that they provide a better user experience. This involves uncovering the problems in the current interfaces for both LaBB-CAT and ElicitSpeech through user studies, create or reorganise information to tailor to users needs, and produce a final prototype that is based on user research. The final prototype will then be benchmarked against the current version in terms of efficiency and effectiveness. Currently for ElicitSpeech, we are looking for the clearest or most efficient design that layman users can easily use and enjoy that can be later adapted for use by people with Dysarthria.

1.3 Scope

This thesis project will cover the requirements gathering process, followed by 2 to 3 phases of iterations of Design, Prototyping, and Evaluation. Target audience would be the main participant pool of the applications, which are naive users for ElicitSpeech, and Speech and Linguistics researchers for LaBB-CAT.

Although more user studies, redesigns and implementation might be carried out in the future, the thesis will only report segments carried out in this thesis project. Therefore, no working website or mobile application will be produced as part of this thesis project. The deliverables will be limited to a paper-based low fidelity prototype, and a high-fidelity prototype built in Axure in the form of sitemaps and wireframes. Detailed aspects of the visual design will not be covered, but the final prototype will be visually designed to be modernized and visually appealing.

I will also be evaluating the user centred approach used in this project, describing my experience with adopting the approach, and my suggestions for future researchers adopting this approach.

1.4 Structure of Thesis

The structure of the thesis is as follows. In Chapter 2 the theories of User-centred design, user experience, background of information architecture, and background of gamification is described. The requirements gathering process for LaBB-CAT is covered in Chapter 3. Chapters 4 and 5 describes the two iterations of the Design, Prototyping, and Evaluation used in the redesign process.

In Chapter 6, the requirements gathered for ElicitSpeech is presented. The three iterations of the Design, Prototyping, and Evaluation used in the redesign of ElicitSpeech are given in Chapters 7, 8, and 9. The conclusion of this thesis follows in Chapter 10.

Chapter II

Theory

In this chapter, we begin with a review of the theory of User-Centred Design and User Experience in Sections 2.1 and 2.2, followed by a description of Information Architecture (Section 2.3) and Gamification (Section 2.4) theories.

2.1 *User-Centred Design*

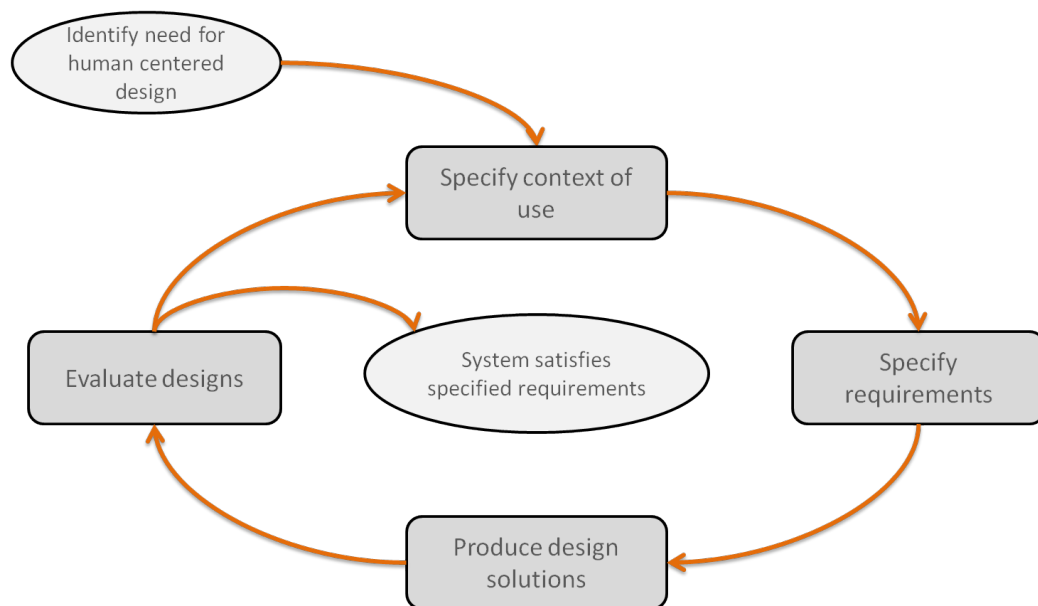


Figure 2.1: General process for user centred design throughout a development life-cycle [1]

User-centred design is a design approach or method that prioritises the users needs, and involves them in the design process. It is distinguished by the active involvement of users and a distinct apprehension of user and task

requirements, a distribution of function between technology and users, repeatedly testing and refining design solutions, and having a multi-disciplinary design [2].

As shown in Figure 2.1, there are four main activities in the process for a user-centred design. First, to specify the context of use, the target audience of the system will need to be identified, and the conditions and uses for the system will need to be found. Second, the business requirements and user goals must be specified. Third, the design solutions will need to be produced, starting from a raw concept to a complete design. Lastly and most importantly, the designs are evaluated through usability testing with users. When the system is found to be successfully satisfying requirements, the product can then be released and the process ends.

2.2 User Experience



Figure 2.2: Honeycomb model of User Experience [27]

The main goal of using a user-centred design approach is to improve and provide a good user experience. This encompasses not only usability, but also

the users perception and emotions towards the system that resulted from its use or anticipated use [27]. Morville [27] suggests a honeycomb model to describe user experience, as shown in Figure 2.2.

Morville [27] explains that every facet of the honeycomb is necessary, but not sufficient to engineer a good user experience. He further believes that there is more to user experience than the 7 facets drawn in the honeycomb model, and that we can perhaps discover and add new boxes to this model in the future [27]. According to Garrett [21], there are many elements to user experience, and there is also a difference in the elements of user experience for a software interface and a hypertext system (Figure 2.3).

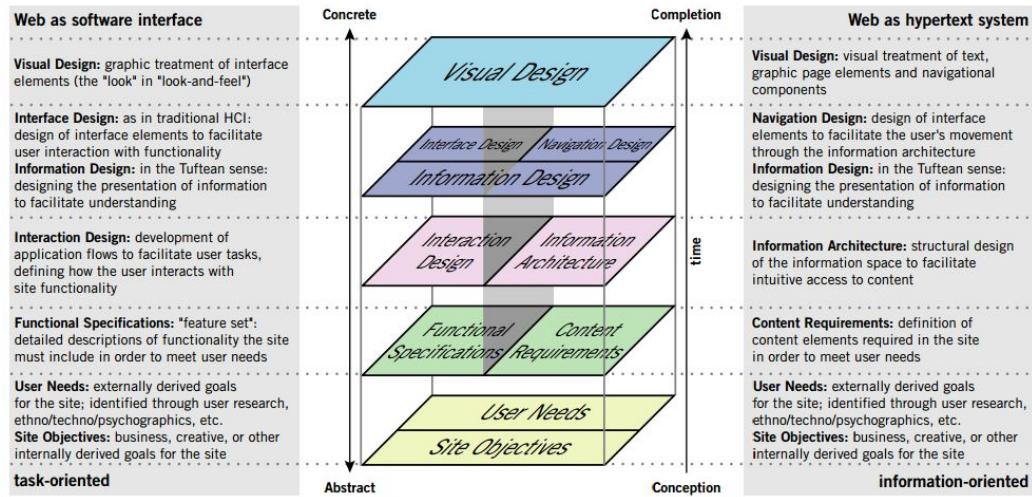


Figure 2.3: Layers of User Experience [21]

There are a few ways to improve user experience as shown in , two which are used in this project are changes to the Information Architecture of LaBB-CAT and the incorporation of Gamification elements into ElicitSpeech. These two aspects are studied in closer detail below.

2.3 Information architecture

Information architecture is the structural organisation of the information or content in the system. It was defined by the Information Architecture Institute as "the practice of deciding how to arrange the parts of something

to be understandable” [9]. Information architecture was first brought up by Richard Saul Wurman in 1975 as an important part of user experience design. Following the boom of the internet and electronics age, information architecture became a popular research topic. Many organisations see information architecture as an important tool for improving usability and findability. Through improving usability, Information Architecture has attracted much interest through e-business and e-government initiatives

A well-executed IA helps people to understand the interface and makes it easy to find, navigate and achieve their goals [9]. If a consumer is accessing a site for the first time, he should be able to quickly understand it effortlessly. By being able to quickly reach their goals, users are able to avoid wasting time and money. This makes IA one of the most cost effective ways to achieve customer satisfaction. Figure 2.4 shows the interdependency of users, content and context, which forms the basis of a good IA design. The main parts are the organization, labelling, navigation, and search systems. Changing a system’s information architecture can improve its usability.

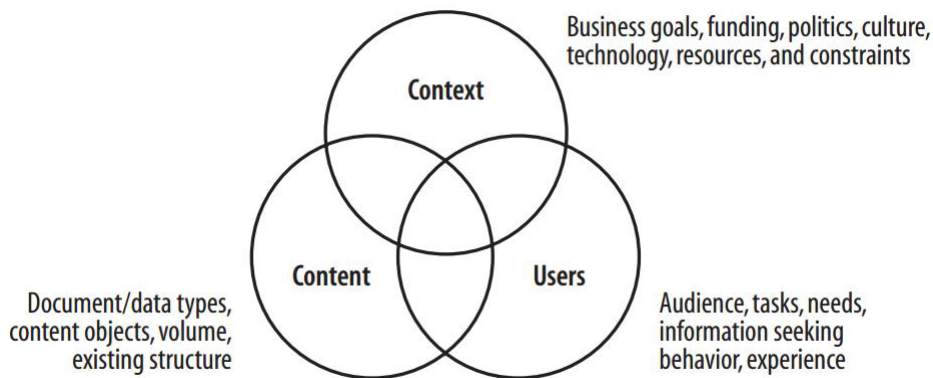


Figure 2.4: Venn diagram of the three circles of information architecture [27]

Context is unique to each organization. The IA designer must understand the business context by understanding questions such as the uniqueness, strategy and future goals of the business. It is also important to perceive the constraints caused by the technology which facilitates interaction between the user and the organisation.

Content is basically what the system is made up of. According to Rosen-

feld [33], content can be further branched out into the following factors: structure, ownership, format, metadata, volume and dynamism (Figure 2.5). All these factors allow for a degree of freedom in customisation.

Users refer to the people who will be using the product. The aim is to study the users and analyse the differences in their preferences and behaviours. It is important to know the importance of users and to make efforts to understand their demographics, aesthetic preferences and purchasing behaviours [33]. By doing so, layouts could be changed and to build upon a niche market different from competitors.

Rosenfield and Morville [28] share that IA can be broken down to four components: Organisation systems, labelling systems, navigation systems and searching systems. Organisation systems are ways which content can be categorised, and is how the relationships between pieces of content are defined or connected. These can be further branched into schemes and structures.

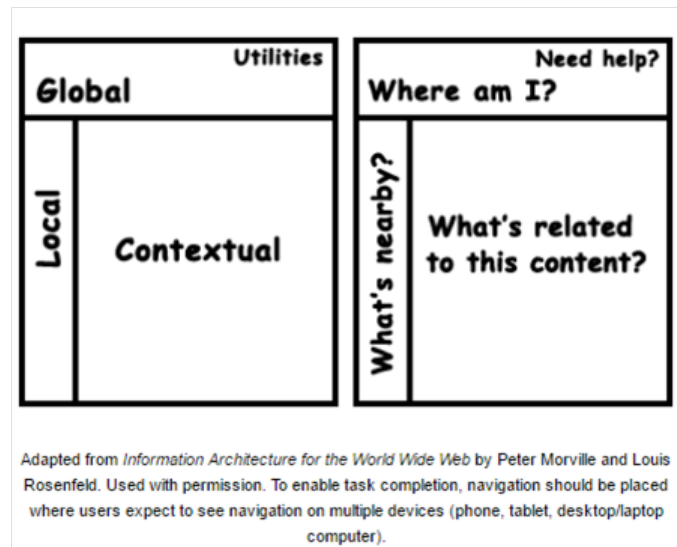


Figure 2.5: Navigation design [12]

For information architecture in any system to be successful, it has to combine user needs, business context, and the changing content. A method suggested by Morville & Rosenfield [28] details the process in five steps. First, research allows a thorough understanding of the business context, content, and the target users through literature reviews and requirements gathering.

Next, a high-level framework strategy is created based on the research to guide the design. The main information architecture is then detailed by deliverables such as wireframe designs. Fourth, the website is implemented by building, testing, and launching based on the design specified. Finally, the existing information architecture is maintained or improved when content is updated in the website. In this project, the first to fourth steps would be followed and reiterated where necessary.

2.4 Gamification

Gamification has been defined as the use of elements characteristic to game design in non-game contexts [16], or the process of augmenting services to create gameful experiences [23]. These are used to elicit a desired type of behaviour in users [18], such as to prolong the use of the application or promote the purchase of in-game items. Nicholson [29] believes that meaningful gamification has to put users at the center of the project, and that it should incorporate user-centred game design elements instead.

All the definitions above mention using “elements of game” instead of turning the entire product into a game. According to [29], one common way to do this is to incorporate scoring elements such as levels, points, or achievements, and apply them to the product to motivate the behaviours desired [15]. Other game elements include but are not limited to achievements, status, combos, bonuses, quests, and levels. These have been proven to work well in real-life implementations and promote uses of commercial software and applications such as IBM’s Beehive system [23], Foursquare [15], and Nike+ [15]. These gamification methods increase uses of services and change user behaviour, as users attempt to satisfy the goals to obtain rewards [29].

Some game mechanic elements currently used in the game industry are shown in Figure 2.6 [24, 35]. Other game mechanics identified by [26] are computerized clues, espionage, pervasive gaming, superimposed game world, secret partnerships, body-mapped avatar, player-undecidable conditions, encouraged face-to-face information exchange, implicit player input, and many others.

Prins et. al [32] found that incorporating game elements into training

1. Achievement	2. Appointment dynamic	3. Avoidance
4. Behavioural contrast	5. Behavioural momentum	6. Blissful productivity
7. Cascading information theory	9. Communal discovery	10. Companion gaming
11. Contingency	12. Countdown	13. Cross situational leader-boards
14. Disincentives	15. Endless games	16. Envy
17. Epic meaning	18. Extinction	19. Fixed interval reward schedules
20. Fixed ratio reward schedule	21. Free lunch	22. Fun once, fun always
23. Interval reward schedules	24. Lottery	25. Loyalty
26. Meta game	27. Micro leader-boards	28. Modifiers
29. Moral hazard of game play	30. Ownership	31. Pride
32. Privacy	33. Progression dynamic	34. Ratio reward schedules
35. Real-time v. delayed mechanics	36. Reinforcer	37. Response
38. Reward schedules	39. Rolling physical goods	40. Shell game
41. Social fabric of games	42. Status	43. Urgent optimism
44. Variable interval reward schedules	45. Variable ratio reward schedule	46. Viral game mechanics
47. Virtual items		

Figure 2.6: Table of game mechanic elements [24, 35]

programme significantly improved motivation and performance of children with ADHD. On the other hand, [13] found that graphics, sound, and storyline are important aspects to have in games, and that the perceived skills required to play the game such as visualisation, logic, and memory are also as important to entice players.

Zichermann & Cunningham [37] believe that gamification drives engagement by increasing user motivation. They highlight the importance of understanding user motivation to build successfully incorporate gamification into systems. There are two forms of motivation: Intrinsic and Extrinsic [37]. Intrinsic motivations manifest from within us and are not affected by our surroundings [37]. On the other hand, extrinsic motivations are driven by our surroundings [37]. According to [37], instead of trying to align extrinsic and intrinsic motivations, they believe that designers should accept user motivational states as they are and push them towards the direction they would like to go and the place we want them to be. They also believe that extrinsic motivators are important to help users discover their intrinsic motivations [37].

Two ways that gamification could increase motivation is through the use of reinforcements or the creation of flow. Flow is the state of losing track of space and time, and is the main aim of designers looking to improve engagement [37]. As shown in Figure 2.7, the tasks designed must neither be too challenging nor too boring so that players will stay in the Flow Zone

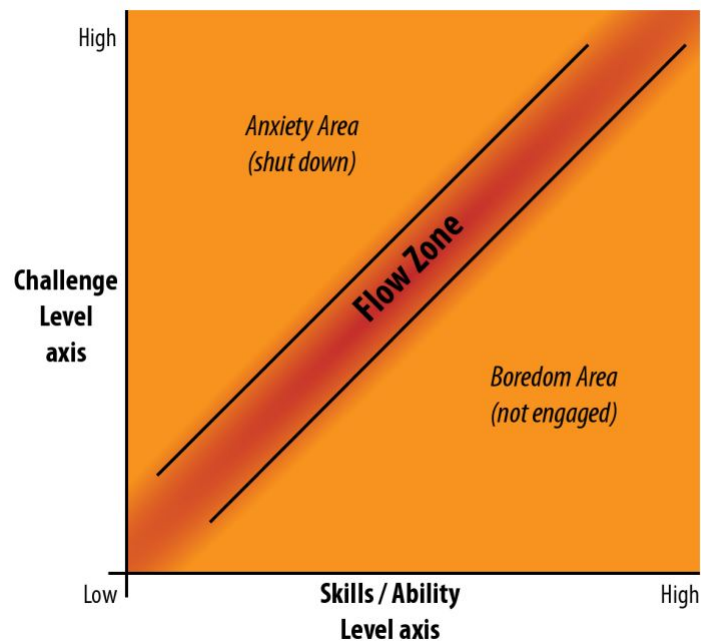


Figure 2.7: Zone where flow occurs [37]

and remain engaged.

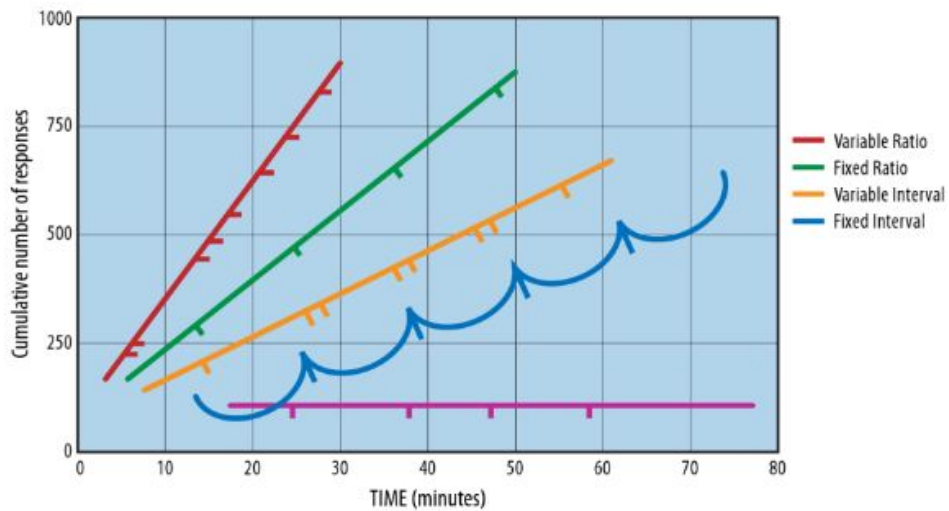


Figure 2.8: Graph of varying kinds of intervals on effectiveness of reinforcement as a motivator [37]

Reinforcement is another way to increase motivation. Many famous psy-

chology studies on reinforcement popularized by Ferster and Skinner ([17]) shows that positive reinforcements can be used to control animal and human behaviour. The intervals at which reinforcements are provided are critical to the success of the reinforcement as a motivator. As seen in Figure 2.8 below, reinforcements provided at variable ratios is the most effective method for increasing motivation.

Zichermann & Cunningham [37] further believe that designers need to know their users to create a better game design. This is in line with the Nicholson's [29] believe that the user needs to be in the centre of the gamification process to make it meaningful.

2.5 Summary

This chapter defined some of the terminology used in the field of user-centred design. Two different systems will be redesigned using a user-centred design approach in this masters project. The different phases used in the redesign will be covered in detail in the next few chapters.

Although there are many methods to improve user experience, the main focus of this thesis aims to bring about improved user experience through the restructuring of information architecture and integrate gamification into the systems depending on their needs.

Chapter III

LaBB-CAT Requirements Gathering

The following three chapters will cover the redesign process of LaBB-CAT. The process of requirements gathering for LaBB-CAT will be detailed in this chapter. The next two chapters will cover the two phases of the redesign, with each phase consisting of a design, prototyping, and evaluation component.

In this chapter the background and current interface of LaBB-CAT are detailed in Sections 3.1 and 3.2. The methodology for the usability testing performed is covered in Section 3.3, and the results obtained from the requirements gathering phase is described in Section 3.4. A summary of the requirements gathering process for LaBB-CAT is presented in Section 3.5.

3.1 Background of LaBB-CAT

LaBB-CAT is a database for media recordings and their time-aligned text transcripts. Developed by the University of Canterbury, it is an interactive digital repository that houses a large audio corpus on a central server. The media files are tagged for participant attributes, transcript attributes, and layers of annotation, allowing researchers to search and interact with the sound files easily.

The search function provides links to the interactive transcript page, allowing users to see the words that match within the context of the entire transcript. The features of the interactive transcript page include but are not limited to viewing of the transcript text, playing back the media, opening it with Praat for further analysis, and selecting annotation layers to display. This allows the acoustic analysis to occur directly through the LaBB-CAT interface. Results of the search can also be exported into Excel, together with

the hypertext links to the relevant sound files in the interactive transcript page.

The browser based version of LaBB-CAT makes it convenient and easily accessible even without installation of software. The graphical user interface makes it easy for non-experts to use, and the search function creates an easy way to search for expressions and texts. LaBB-CAT is created to be a convenient tool for linguistics research. The results can be exported into Microsoft Excel with the hyperlinks to its sound files. It also allows researchers to be able to access information of audio, the speakers, as well as the transcript with layered annotations all in one website.

NZILBB is currently using one of the LaBB-CAT servers to study dysarthria media recordings. This study would consist of two components. Participants would be recruited to take part in a speech recording exercise using ElicitSpeech, where they would be asked to speak a set of sentences into a microphone, either using an iPad or the web browser. These people are henceforth referred to as speakers. Each speaker should only be allowed to take part in the exercise once. In the meantime, researchers use LaBB-CAT to study the data and audio recordings obtained through ElicitSpeech.

In this study, the focus is only to redesign the interface of the LaBB-CAT instance and the associated application that is being used for the NZILBB dysarthria study. This comprises of:

1. LaBB-CAT: Browser-based interface used by researchers of the NZILBB study
2. ElicitSpeech: Browser- & Tablet- friendly Interface used by speakers

3.2 Current web interface of LaBB-CAT

Screenshots are included in this section that shows a quick overview of the main pages of LaBB-CAT. The current web interface is a very plain and basic interface, where the essential words are shown as links. It is a mix between a software interface and a hypertext system.

In my first use of the platform, I found it difficult to learn. Although the design was simple as shown in Figure 3.1, but I was unsure of the features

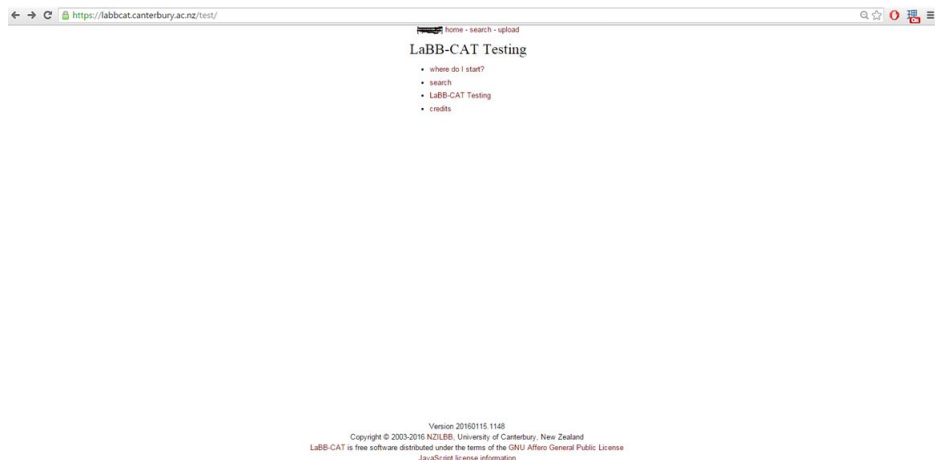


Figure 3.1: Home page of current LaBB-CAT platform for researchers (web)

that the application could offer, and had to spend some time figuring that out. There was too much white space, too much text, and little to no images. I also did not understand the home page of the LaBB-CAT, as there was a heading “LaBB-CAT Testing as well as a navigational link with the same title, so I was not sure where that link would go. It would be better to rename the link as “LaBB-CAT test statistics, and the heading could be renamed “LaBB-CAT for easier understanding.

The version number, copyright, and license information was also displayed in the bottom centre of every page, which I did not like. Perhaps it can be either removed or moved to the side so that it is less prominent.

I was also uncomfortable with the search page (Figure 3.2), as I was swamped with results I did not understand. The results were also not aligned according to the filter function. When one filter was applied, the results narrowed down to a single one, which was not to my expectations. Moreover, the results that surfaced were not understandable to me, and I had difficulty understanding how to use the search function. As I was unfamiliar with speech and language research terminology, it is possible that these problems surfaced due to my inexperience in the field.

The screenshot shows a web browser window with the URL <https://labcat.canterbury.ac.nz/test/filter>. The page displays a search filter interface with a table of results. The table has columns: name, transcripts, layers, corpus, Gender, and Birth Year. The data is filtered to show records where the corpus is 'en' and the gender is 'Male'.

name	transcripts	layers	corpus	Gender	Birth Year
test-108	[R-0]	[R-0]	en	Male	123
test-122	[R-0]	[R-0]	en	Male	123
test-128	[R-0]	[R-0]	en	Male	123
test-102	[R-1]	[R-1]	en	Male	1972
test-103	[R-1]	[R-1]	en	Male	123
test-104	[R-1]	[R-1]	en	Male	111
test-105	[R-1]	[R-1]	en	Male	111
test-107	[R-1]	[R-1]	en	Male	123
test-109	[R-1]	[R-1]	en	Male	111
test-110	[R-1]	[R-1]	en	Male	111
test-111	[R-1]	[R-1]	en	Male	111
test-112	[R-1]	[R-1]	en	Male	123
test-113	[R-1]	[R-1]	en	Male	123
test-114	[R-1]	[R-1]	en	Male	123
test-115	[R-1]	[R-1]	en	Male	123
test-116	[R-1]	[R-1]	en	Male	123
test-117	[R-1]	[R-1]	en	Male	123
test-118	[R-1]	[R-1]	en	Male	123
test-119	[R-1]	[R-1]	en	Male	123
test-120	[R-1]	[R-1]	en	Male	123
test-121	[R-1]	[R-1]	en	Male	123
test-123	[R-1]	[R-1]	en	Male	123
test-124	[R-1]	[R-1]	en	Male	123
test-98	[R-1]	[R-1]	en	Female	1991
test-106	[R-2]	[R-2]	en	Male	111
Atax_F15_Chile	[R-3]	[R-3]	Chile_Fieldwork_Dysarthric		
Hunt_F13_Chile	[R-4]	[R-4]	chile_fieldwork_dysarthric		
Hunt_F14_Chile	[R-4]	[R-4]	chile_fieldwork_dysarthric		
Hunt_F15_Chile	[R-4]	[R-4]	chile_fieldwork_dysarthric		
LCVA_F17_Chile	[R-4]	[R-4]	chile_fieldwork_dysarthric		
LCVA_F21_Chile	[R-4]	[R-4]	chile_fieldwork_dysarthric		
LCVA_F23_Chile	[R-4]	[R-4]	chile_fieldwork_dysarthric		
Pak_F19_Chile	[R-4]	[R-4]	chile_fieldwork_dysarthric		
Pak_F22_Chile	[R-4]	[R-4]	Chile_Fieldwork_Dysarthric		
Pak_F23_Chile	[R-4]	[R-4]	Chile_Fieldwork_Dysarthric		
Pak_F25_Chile	[R-4]	[R-4]	chile_fieldwork_dysarthric		
Pak_F27_Chile	[R-4]	[R-4]	chile_fieldwork_dysarthric		
Canb_F8_Chile	[R-5]	[R-5]	chile_fieldwork_dysarthric		
CVA_F5_Chile	[R-5]	[R-5]	chile_fieldwork_dysarthric		
Gullabarr_F20	[R-5]	[R-5]	chile_fieldwork_dysarthric		
Hunt_F12_Chile	[R-5]	[R-5]	chile_fieldwork_dysarthric		
Hunt_F15_Chile	[R-5]	[R-5]	chile_fieldwork_dysarthric		

Figure 3.2: Search page of current LaBB-CAT platform for researchers (web)

3.3 Usability testing of LaBB-CAT's current interface

Gathering requirements is one of the most essential elements of a User-Centred Design methodology. It includes gathering of feedback, surfacing of problems, as well as finding out the essentials missing from the current interface. This usually occurs at the beginning phases of any design project, and requires the involvement of target users.

The target users of LaBB-CAT are linguistics researchers. LaBB-CAT was designed as an interface to allow linguistics researchers easy access to search the corpus, listen to the recordings of speakers, and extract the data.

3.3.1 Participants

Eight speech and linguistics students at the University of Canterbury participated in the user study. Of the participants, 87.5% were females and 12.5% were males. The mean age of the participants was 28.25 with a standard deviation of 4.13 in a range of 22 to 35 years old. The participants were recruited through posters placed in student advertisement locations all throughout the university, as well as by contacting linguistic students who have previously used LaBB-CAT for their research.

3.3.2 Design and Apparatus

This study was conducted as requirements gathering through usability testing of the current interface. The dependent variables were the qualitative feedback obtained in the semi-structured interview.

The current LaBB-CAT interface was used in this study. The participants were asked to complete four different tasks on LaBB-CAT. These tasks were obtained from the application developer and were representative of tasks that researchers commonly perform on LaBB-CAT. The tasks were displayed one per page to ensure minimal distractions. A definitions sheet was placed on the table with the definitions of the terminology used in LaBB-CAT for easy reference whenever required.

3.3.3 Procedure

The study was conducted in a quiet and bright room in the University of Canterbury. Before the study began, participants were read an introduction paragraph ensuring them that the study is evaluating the interface and not them. It was specifically mentioned that there are not right or wrong answers to any of the questions asked in the study, and that they can feel free to speak their mind for the benefit of the product. They were briefed about being video recorded, and that they should think aloud throughout the entire session. They were then given the information sheet and consent form to sign.

A semi-structured background interview adapted from Russell [34], a Google Researcher, was then conducted where they were asked about their previous experience with LaBB-CAT and other linguistics or speech research databases (Appendix A). Other background information was also collected such as their field of study or research, age, gender, work environment.

They were then instructed to complete the tasks on a LaBB-CAT instance provided using the desktop. Upon completion of each task, they were asked questions specifically targeted to find out problems they had encountered in the specific task. After completion of all four tasks, they were given another semi-structured interview, which consisted of open-ended questions adapted from the System Usability Survey (SUS) [7, 4] as well as questions that allowed them to voice their opinions regarding what they liked and disliked

about the interface and their latent desires for tasks. Upon completion of the study, they were awarded a \$20 mall voucher.

3.4 Results

Four of the participants were majoring in Speech and Language therapy and the others (50%) were doing Linguistics postgraduate research. Of the 8 participants, 6 had experience with using LaBB-CAT, but only 1 of them felt they were experienced in LaBB-CAT, while the rest of them responded that they were just average or not experienced. Out of the 6 participants who had experience with using LaBB-CAT, 5 of them used LaBB-CAT for searching for various data to export, and all 6 of them (100%) mentioned that they required the help of either their supervisor or the software developer to complete the tasks. The amount of time that most participants use LaBB-CAT per week is based on their needs. They could use up to 3 hours per day when required for their research, or as little as 0 hours per week.

From observation of participants completing the tasks given in the user study, it was found that most participants are not able to search and export concisely the data that they want. They would either obtain data sets that was incomplete and conduct another search and export for the missing data, or they would obtain data sets that contained too many of the data that they do not require. Most participants were also not confident in their results.

It was also found that all of the participants (100%) more than 5 minutes to find the Layered Search, and most participants (62.5%) needed help to find it. After accessing the Layered Search, more than 75% of participants were unsure of how it works, and ended up with incorrect results.

From the feedback section, a few participants mentioned that they were frustrated about the loss of data when they had created an incorrect search. Although the interface was difficult to use, participants mentioned that they could always rely on the developer to provide them with the help they require, but this would be highly taxing and costly for the developer.

3.5 Summary

From the data gathered from the user study, the results were consistent with initial observations that the interface is difficult to use and learn. It was not user-friendly and even experienced users of LaBB-CAT are unable to understand how to use the interface to complete the most basic and common tasks without help. All experienced participants mentioned that they almost always required help from either their supervisors or the software developer, which is time-consuming for all involved.

From this requirements gathering, it was decided that the interface required a complete redesign from the way the information is presented to the users, to the way users can search for the data they need. The main requirements for this interface listed in order of importance are shown below:

1. Users need to be able to conduct searches easily, confidently, accurately, and efficiently.
2. Layered Search needs to be accessible and usable
3. Export needs to be more efficient
4. Transcript needs to be effective, efficient and satisfactory
5. Navigation bar needs to be consolidated
6. Feature requests
7. Help needs to be more specific to the target area

Chapter IV

LaBB-CAT Redesign Phase 1

This chapter will cover the first iteration of Design (Section 4.1), Prototype (Section 4.2), and Evaluation (Section 4.3). The design was created based on the six main requirements found in the previous section, namely problems with Search, Layered Search, Export, Transcript, Navigation bar, and Help. Results found from the evaluation are described in Section 4.4 and a summary of Phase 1 of the redesign for LaBB-CAT is covered in Section 4.5.

4.1 Design

Many of the main features in LaBB-CAT were redesigned to improve usability. First, the Search was completely changed to follow similar popular database search designs such as iTunes and EndNote. These software contain a huge database and allow their users to search through their collection quickly and easily. By following their concept of adding search criteria and using several dropdown lists, “And/Or/Not” functions, the new LaBB-CAT search function was created. As there was only two main categories of data that can be searched and displayed, namely the Participant information and the Transcript layers (Figure 4.1), the Search now shows the clear distinction between the two.

The Layered Search was found to be confusing and difficult to access. It was thus converted to an Advanced search, where users could use to search across any number of words they need. A button label for “Search across 2 or more words” was also placed in a prominent location on the Search page for easy access.

Search and Export only allowed one set of data to be displayed and export

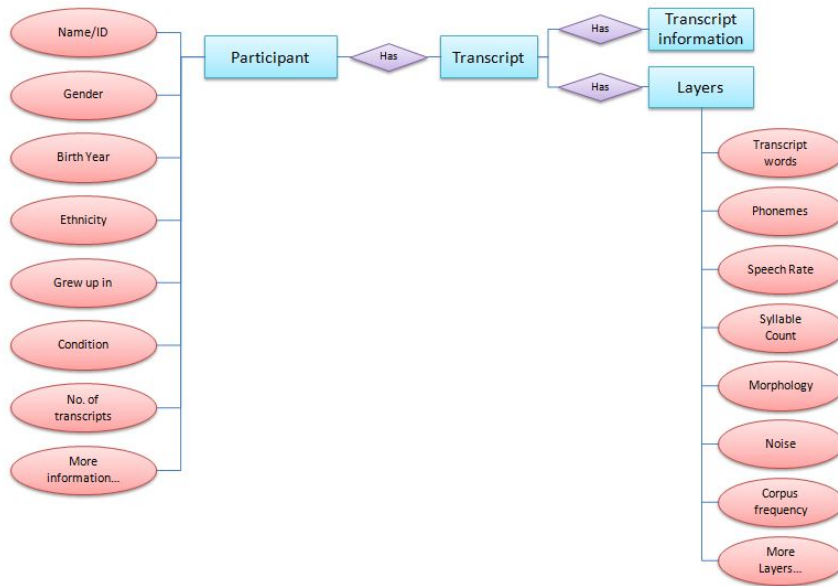


Figure 4.1: Data organisational structure of LaBB-CAT

at any one time, and users had to manually search for additional data sets that they wanted or remove the data sets that were not required, which is time costly. This was improved by creating the “Search Criteria”, which would allow users to search two or more different sets of data and displaying them on a single page easily and neatly.

Although the Transcript is one of the main information that users need to access, the old interface required the navigation of up to 6 pages reach it. The structure of page navigation was changed so that users can now access the data within two navigational clicks. The participant information for each transcript was also combined into the same page as the transcripts display for easy referral. The transcript display was also redesigned to be less confusing. The layers that can be activated on the transcript was placed on the right hand side so that users can easily select the layers that they want displayed on the transcript, and participant information can be easily seen on the left hand side, with the option to be minimized to reduce clutter.

The navigation bar was one of the easier features to redesign. The old interface had confusing different navigation links in the home page and on

the top of every page. These were combined into a single navigation bar on the top of every page, which had links to all the features available in LaBB-CAT. Moreover, the home page with links was replaced by the Search, as it is the feature that is used the most in LaBB-CAT.

The Help function which was too lengthy and often unread by users was replaced by a video tutorial, which is displayed in the paper prototype as a live demonstration of important features in the page. Further, small question mark symbols are also attached to important widgets, which would allow users to obtain information specific to the widgets.

Several features were also added to LaBB-CAT as a result of the requirements gathering process. Following participants feedback regarding loss of data, the Search History feature was created that would allow easy access to their previous searches.

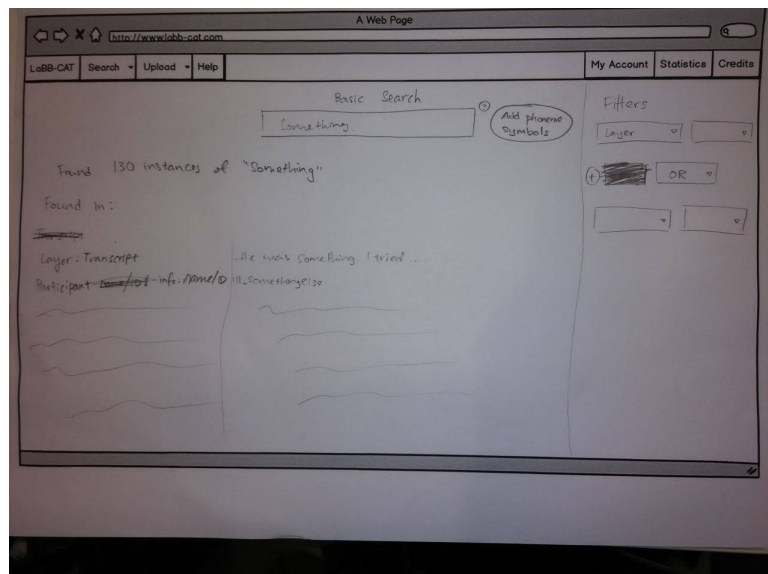


Figure 4.2: Data organisational structure of LaBB-CAT

Several sketches were made as seen in Figures 4.2 and 4.3, and refined over time to be a single prototype.

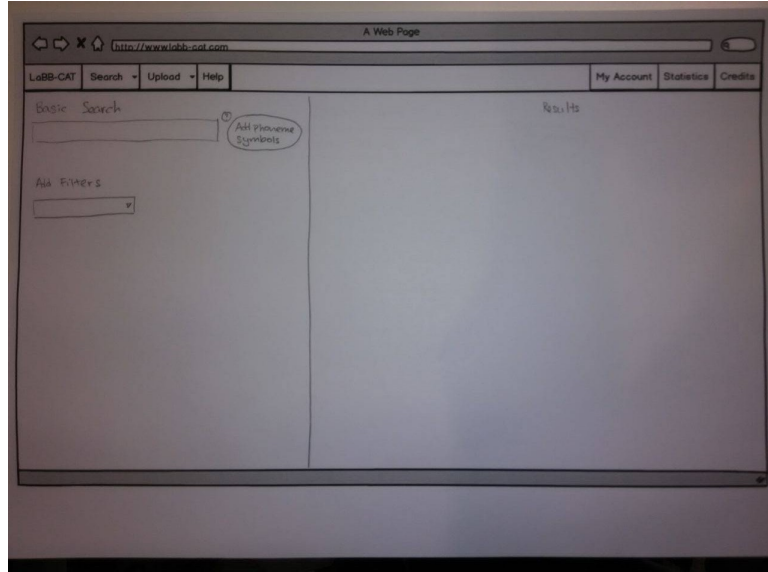


Figure 4.3: Data organisational structure of LaBB-CAT

4.2 Low-fidelity prototyping of LaBB-CAT

Low fidelity (Lo-Fi) prototypes were created in the early phases of the design process, following the requirements gathering user study. These prototypes are easy to create, and provides an easy way to obtain feedback about the design quickly. The low-fidelity prototype was created using a combination of paper and Balsamiq ¹, a rapid wireframing tool.

The browser display, main navigation headers and some simple pages such as participant information and transcript display were created in Balsamiq so that they can be printed and duplicated easily as seen in Figure 4.4. Buttons that do not require interactivity are drawn onto the Balsamiq papers as seen in the Home and Advanced Search page (Figure 4.5).

To add interactivity to the paper prototype to simulate the actual interface, widgets are created as shown in Figure 4.6. These widgets are flexible and reusable, and can be placed on the interface when buttons are pressed and can be moved around the page easily (Figures 4.7, 4.8). Some of the interactions displayed by widgets include:

¹ <https://balsamiq.com/>

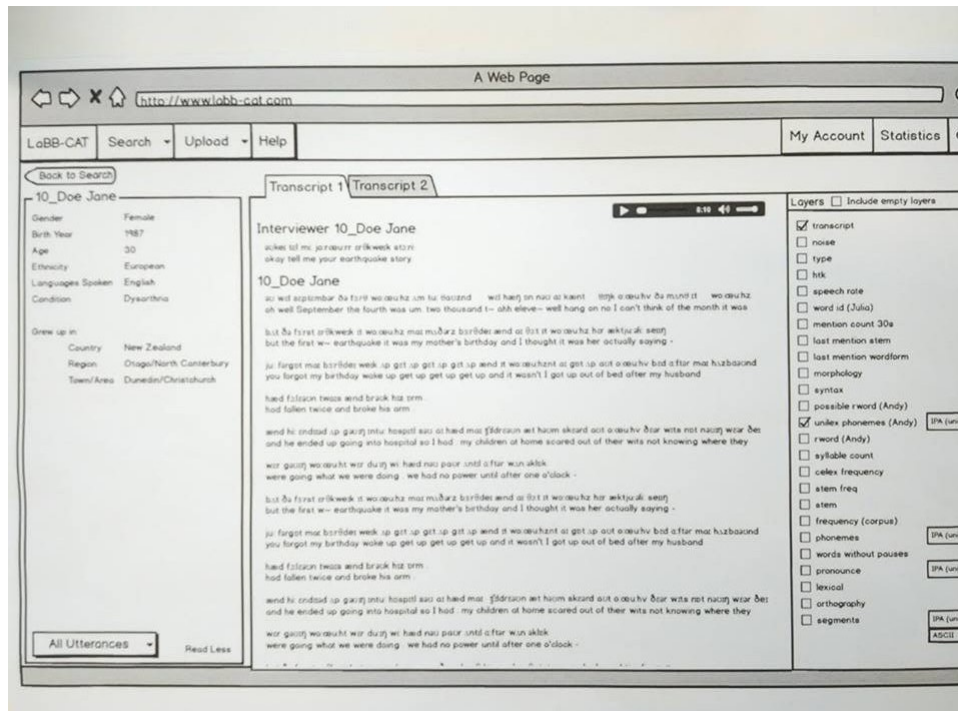


Figure 4.4: Transcripts and Participants information page used in the Lo-Fi Prototype of LaBB-CAT

- Dropdown Lists
- Search boxes
- Error message popups
- Export popup
- Search results
- Plus button
- Additional search criteria
- Symbols help
- Shorthand help

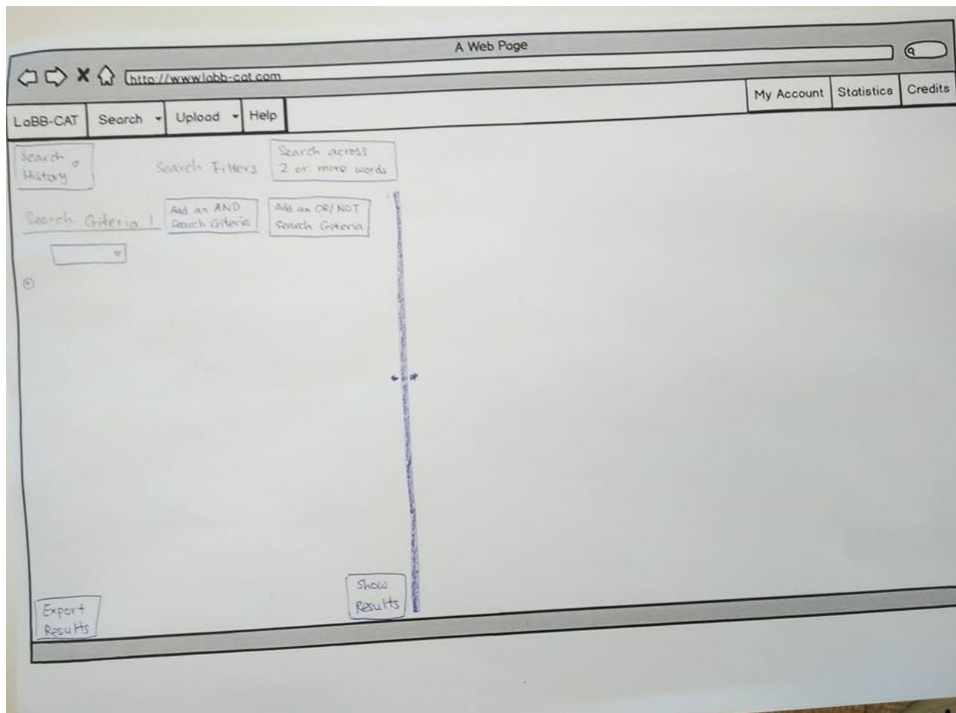


Figure 4.5: Basic Search page used in the Lo-Fi Prototype of LaBB-CAT

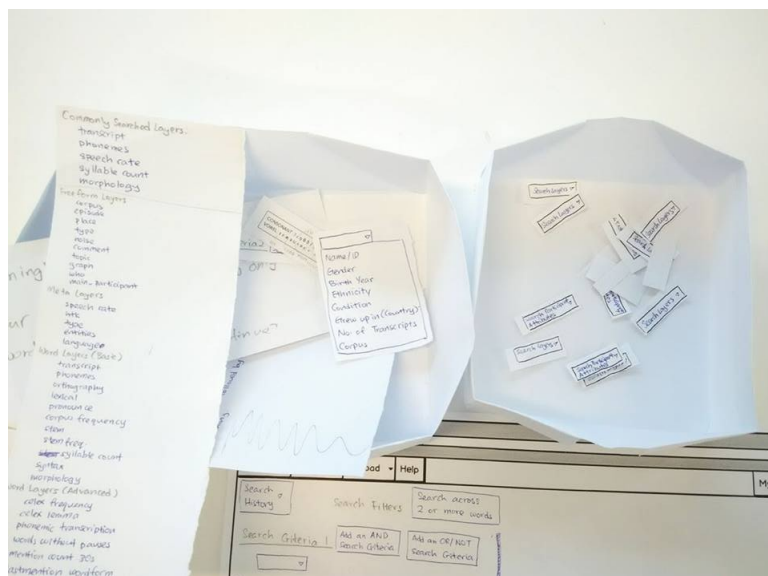


Figure 4.6: Widgets used in the Lo-Fi Prototype of LaBB-CAT

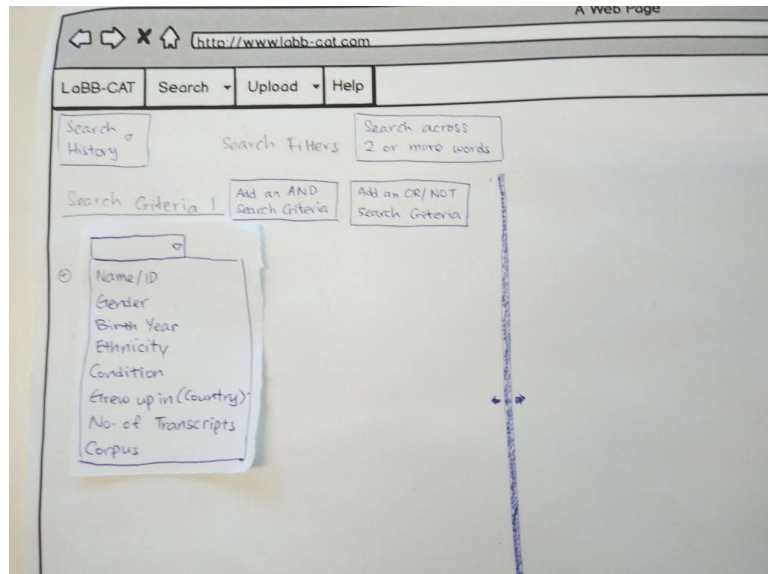


Figure 4.7: A dropdown widget in use in the Lo-Fi Prototype of LaBB-CAT

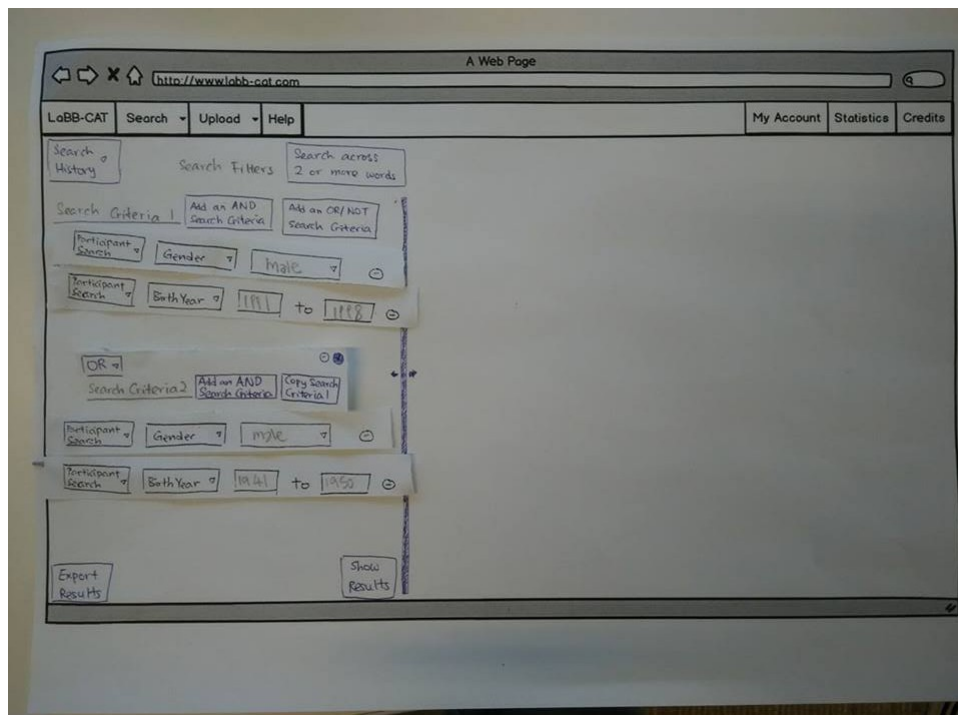


Figure 4.8: A search created in the Basic Search page used in the Lo-Fi Prototype of LaBB-CAT

4.3 Usability testing of LaBB-CAT's low-fidelity prototype

Usability testing is one of the most important steps in a user-centred design. It tests the prototype against the target users to obtain results and ideas for design improvement. It can be used in the early phases to ensure the prototype is sufficiently user-friendly before it is advanced onto the next phases of the redesign, to ensure the costs of the project is kept within budget and on time.

4.3.1 Participants

Nine speech and linguistics students at the University of Canterbury participated in the user study, five of whom (55.6%) had previously participated in my requirements gathering usability testing. All of the participants (100%) were females. The mean age of the participants was 26.56 with a standard deviation of 4.39 in a range of 21 to 33 years old. The participants were recruited through posters placed in student advertisement locations all throughout the university, contacting speech or linguistic students through their departments, and contacting participants who had previously completed the first user study.

4.3.2 Design and Apparatus

This study was conducted as a Wizard of Oz style usability testing of the Lo-Fi prototype. The dependent variables are the qualitative feedback obtained in the semi-structured interview and the quantitative SUS scores.

The Lo-Fi LaBB-CAT prototype was used in this study. The participants were asked to complete four different tasks on LaBB-CAT. These tasks were similar to those used in the first usability testing, but with some modified steps and numbers. As the prototype designed was significantly different from the old interface, and the results from the first usability testing can no longer be extrapolated to the new prototype, similar tasks were used to find out if there are similar problems in the new prototype. The tasks were displayed one per page to ensure minimal distractions.

4.3.3 Procedure

The study was conducted in a quiet and bright room in the University of Canterbury. Before the study began, participants were read an introduction paragraph ensuring them that the study is evaluating the interface and not them. It was specifically mentioned that there are not right or wrong answers to any of the questions asked in the study, and that they can feel free to speak their mind for the benefit of the product. They were briefed about being video recorded, and that they should think aloud throughout the entire session. They were then given the information sheet and consent form to sign.

A semi-structured background interview was then conducted using a survey software, Qualtrics ², where they were asked open-ended about their previous experience with LaBB-CAT and other speech or linguistic research databases. Other background information was also collected such as their field of study, age, gender, and programming experience.

They were then instructed to complete the tasks on the LaBB-CAT prototype presented on paper. They were told that they could press anything they wish and that they should feel free to explore the interface features. When participants pressed any button on the interface, I would perform an action, either by placing a widget on the page, changing the page to another, or if the action was not predicted or out of scope, I would provide the participants with a description of the intended feature or result of the pressed button to give them an idea of the what it does. When participants pressed the Help at any time, they were provided by the experimenter with a live “video tutorial” describing the main features of the page.

Upon completion of each task, they were asked questions specifically targeted to find out problems they had encountered in the specific task. These included “What was the main question you had with this task?”, and they would be asked to elaborate and explain their answer. After completion of all four tasks, they were given another semi-structured interview (Appendix B), which consisted of open-ended questions adapted from the System Usability Survey (SUS) as well as questions that allowed them to voice their opinions regarding what they liked and disliked about the interface and their latent

² <https://www.qualtrics.com/>

desires for the system. Upon completion of the study, they were awarded a \$20 mall voucher.

4.4 Results

As the first participant had a difficult time understanding the paper prototype, the widgets of the prototype were changed to small individual pieces of paper, which would allow participants to understand the interface better. As she only managed to complete 2 tasks out of the required four in the span of an hour, her results were removed from the analysis.

The mean SUS score (N=8) obtained for this interface was 82.2 with a standard deviation of 15.38. Due to the low number of participants involved in this study, only descriptive statistics and qualitative data are reported. The SUS score from participants who had experience with using LaBB-CAT (N=4, M=88.1, SD=4.27) was higher than from naive users (N=4, M=76.3, SD=20.97).

A one-sample t-test was also conducted to compare the mean SUS score obtained for this prototype with the average SUS score of 68 as obtained from 500 different studies [4]. The mean SUS score for the new prototype (M=82.2, SD=15.4) was found to be statistically significantly higher than the average score of 68 ($p=0.04$) 4.9.

One-Sample Test						
	Test Value = 68					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
SUSScore	2.609	7	.035	14.18750	1.3308	27.0442

Figure 4.9: One-Sample T-Test of prototype against population average of 68

One of the biggest findings from this user study is the problem with the shorthands. From observation, most of the participants did not know that they had to use shorthands, and all of them were confused about which shorthand to use even when explanation was provided. It was suggested that

examples be provided on the widget to show users how they could be possibly be used, and the results that would be obtained from using the shorthands.

Four participants (50%) took more than 3 minutes to complete the first Search task, and three participants (37%) did not manage to complete the first task correctly after 5 minutes. Only one of these participants clicked the Help to get the live tutorial on how to use the interface, but she was clearly still confused after the tutorial ended.

Another finding from the tasks given was that some of the participants (37.5%) did not know that they had to use the Advanced Search, which they would need to search across a two or more words. This could be due to the Task consisting of two different components, which confused participants as they were unsure how to begin. It could also possibly be that the label for the button was ambiguous about its function.

All of the participants (100%) had little to no problems with accessing the transcript and finding various information.

For the question “What did you like about this system?”, five of the participants (62.5%) mentioned the Search as it allowed them to search multiple sets of data on one page, while the rest of the participants (37.5%) mentioned that they liked the simplicity of the system and that it was easy to use.

One observation was that most participants (75%) appeared to improve in performance and found it easier to use the interface as they progressed. However, this would need to be tested in the next phase to reinforce the finding with data and feedback.

4.5 Summary

Of the requirements specified in the previous chapter, three out of six of them are met in this interface, namely the requirements concerning Export efficiency, Transcript efficiency, effectiveness, and satisfaction, and Navigation bar consolidation.

The mean SUS score of the paper prototype obtained was a grade A [4] and the qualitative data obtained did not show critical problems with the interface. As the feedback obtained was generally positive, I decided that the next phase of the redesign would be to proceed with the design, building,

and evaluation of the high-fidelity prototype. The following problems must be targeted to solve the specific findings in this user study:

- Searching across a few words
- Phoneme and shorthand search
- Learnability of the interface

Chapter V

LaBB-CAT Redesign Phase 2

This chapter will cover the second iteration of Design (Section 5.1), Prototype (Section 5.2), and Evaluation (Section 5.3). Similar to the previous chapter, results found from the evaluation are described in Section 5.4 and a summary of the redesign process for LaBB-CAT is covered in Section 5.5.

5.1 Design

As the SUS score of the paper prototype was an A grade, I decided to proceed with the creation and evaluation of a High-Fidelity prototype.

Initial labels on the paper prototype such as “Search Clause” and “Add a Clause” were not clear, which led to some confusion. These confusing labels were gradually changed in the paper prototype based on feedback collected, which were reflected in the final paper prototype. An Advanced Search was included in the navigational bar of the Axure prototype that will bring users to the page allowing for searching across two or more words.

A warning was added as a feedback for participants to know that they might have made a mistake in their search if their search appears suspicious or contradictory. This would lead users to find alternative solutions for their search, such as using the Shorthand symbols button next to the phonemes search box.

The shorthand help was also redesigned to include not only the descriptions of the symbol, but also the various examples of how the symbol can be used and the results that would be included.

5.2 High-fidelity prototyping of LaBB-CAT

The High Fidelity (Hi-Fi) prototype was generally created in the later phases of the design process, following several rounds of usability testing of the Lo-Fi prototype. These prototypes take longer to create, but encompass more of the features and aesthetics of the final prototype that will be implemented at the end of the redesign project.

The Hi-Fi prototype for this project was created using Axure ¹, a software for prototyping and wireframing.

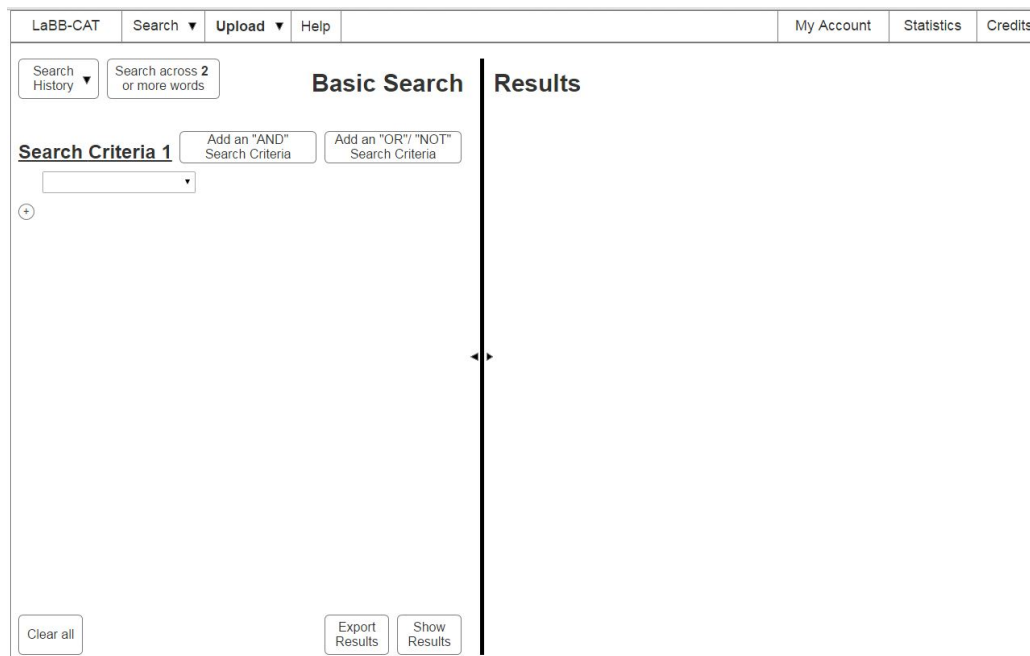


Figure 5.1: Home page of the Axure prototype of LaBB-CAT

5.3 Usability testing of LaBB-CAT's high-fidelity prototype

It was decided that another usability testing would be conducted instead of an experimental evaluation as the qualitative feedback obtained for the old interface was too critical. Furthermore, although this would be the final phase in my master's project, it is not the last phase of the entire redesign

¹ <http://www.axure.com/>

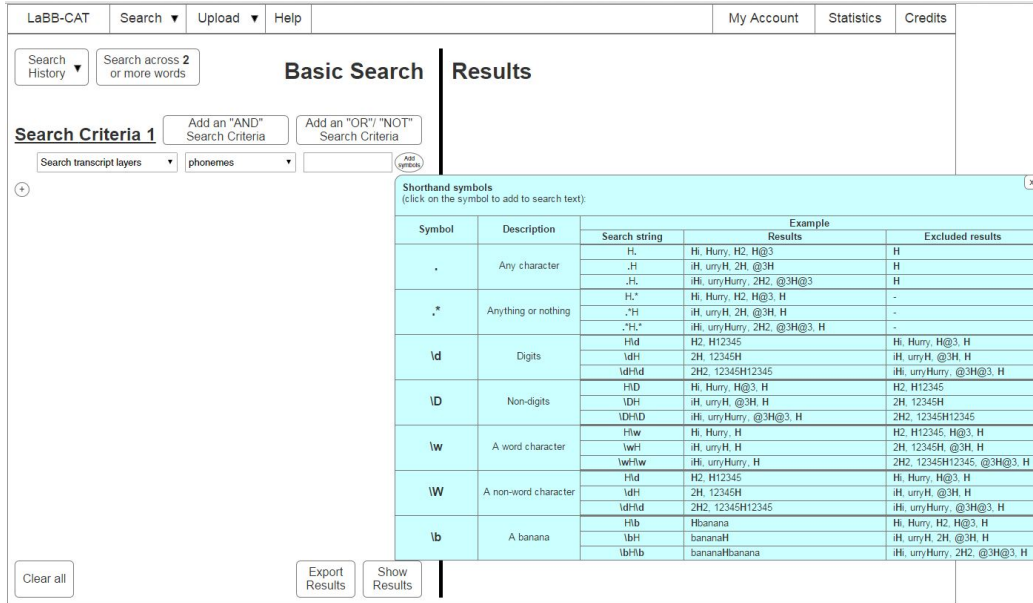


Figure 5.2: Shorthands used in the Axure prototype of LaBB-CAT

project. Results obtained from this usability testing would be much more beneficial for future phases of the redesign project.

5.3.1 Participants

Ten speech and linguistics students at the University of Canterbury participated in the user study. Of the participants, 80% were females and 20% were males. The mean age of the participants was 29.0 with a standard deviation of 4.55 in a range of 20 to 34 years old. Three of the participants had attended both the first and second user studies, 2 participants had each attended one of the previous user studies, and the rest of the participants (5) had not participated in any of my previous user studies. Of the ten participants, 60% had experience using LaBB-CAT, while 40% were naive users of LaBB-CAT.

The participants were recruited through posters placed in student advertisement locations all throughout the university, contacting speech or linguistic students through their departments, or contacting participants who had previously completed the first or second user study.

5.3.2 Design and Apparatus

This study was conducted as a usability testing of the Hi-Fi prototype. The dependent variables are the qualitative feedback obtained in the semi-structured interview.

The Hi-Fi LaBB-CAT prototype was used in this study. The participants were asked to complete four different tasks on LaBB-CAT. The tasks were displayed one per page to ensure minimal distractions. These tasks were created to target specific problems found in the prototype in the previous user study:

- Task 1: Complex Search and Export
- Task 2: Searching across a few words
- Task 3: Phoneme and shorthand search
- Task 4: Learnability of the interface

5.3.3 Procedure

The study was conducted in a quiet and bright room in the University of Canterbury. Before the study began, participants were read an introduction paragraph ensuring them that the study is evaluating the interface and not them. It was specifically mentioned that there are no right or wrong answers to any of the questions asked in the study, and that they can feel free to speak their mind for the benefit of the product. They were briefed about being video recorded, and that they should think aloud throughout the entire session. They were then given the information sheet and consent form to sign.

A semi-structured background interview was then conducted using Qualtrics where they were asked open-ended questions about their previous experience with LaBB-CAT and other speech or linguistic research databases. Other background information was also collected such as their field of study, age, gender, and programming experience.

They were then instructed to complete the tasks on the LaBB-CAT prototype presented on a Google chrome web browser. They were told that they

could press anything they wish and that they should feel free to explore the interface features. When participants pressed any button on the interface that was not predicted or out of scope, I would provide the participants with a description of the intended feature or result of the pressed button to give them an idea of what it does.

Upon completion of each task, they were asked questions specifically targeted to find out problems they had encountered in the specific task. These included “What was the main question you had with this task?”, and asked to elaborate and explain their answer. After completion of all four tasks, they were given another semi-structured interview (Appendix B), which consisted of open-ended questions adapted from the System Usability Survey (SUS) as well as questions that allowed them to voice their opinions regarding what they liked and disliked and their latent desires for the interface. Upon completion of the study, they were awarded a \$20 mall voucher.

5.4 Results

The mean SUS score (N=10) obtained for this interface was 76.0 with a standard deviation of 11.91. As the data was found to be normal, an independent samples t-test was conducted for measuring difference between experienced and naive users. There was a significantly higher ($t=-4.38$, $p=.002$) SUS score from participants who are experienced LaBB-CAT users (N=6, $M=83.8$, $SD=7.54$) than from naive users (N=4, $M=64.4$, $SD=5.54$).

A one-sample t-test was also conducted to compare the mean SUS score obtained for this prototype with the average SUS score of 68 as obtained from 500 different studies [4]. There was no statistically significant difference ($p=0.06$) between the mean SUS score ($M=76.0$, $SD=11.9$) of the new interface as compared to the population score of 68 5.3. However, there appears to be a trend towards significance.

50% of the participants were able to complete the complex search and export task in one try. 30% of the participants completed the task in one try as well, but did not notice that they had not changed the dropdown list of “OR/NOT” to “NOT”. One of these participants eventually realised in Task 4 that she had selected the wrong “OR/NOT” option. All of the

One-Sample Test						
	Test Value = 68					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
SUSScore	2.123	9	.063	8.00000	-.5228	16.5228

Figure 5.3: One-Sample T-Test of new interface against population average of 68

participants who completed the task in one try (80%) were confident in their results. 20% of the participants did not manage to complete the task in one try. One of them was confused in the first try, but managed to successfully complete the search and export in her second try. The other participant who had not managed to complete the task in one try used the Help to get a live tutorial of the basic feature of the Basic Search, followed by an example of a Search. She easily and quickly completed the complete search and export task after the tutorial.

Only 10% of the participants managed to complete the task to search across two or more words in one try. 70% of the participants had trouble finding out how to search across more than one word, but 5 out of 7 of them managed to complete the task after using the Help for a tutorial. 40% of the participants did not craft the correct search parameters, indicating that they did not completely understand how the “OR” search criteria works.

It was found that most of the participants (90%) were still confused by the shorthands when they used the prototype. Only 20% of the participants managed to complete the task correctly for the shorthand without Help. 10% of the participants completed the task confidently but used the wrong shorthand. 60% of the participants managed to use the correct shorthand after a tutorial, and 10% of the participants did not manage to complete this task even with a tutorial. Although descriptions and examples of shorthand uses were provided in the new interface, these design solutions did not solve the problem and many participants were still confused about how to use them. Future redesign phases should either completely remove the shorthand feature and integrate it naturally into the search, or provide a short video tutorial about what shorthands are and how they work. However, it was

found that most participants (70%) understood the shorthand after having used it once, and could correctly apply it in a slightly different context in the next task.

90% of the participants displayed understanding of the results shown to them, and only 10% of the participants found the results confusing.

80% of the participants completed the last task quickly, easily, and confidently. Of the 8 participants, 5 of them had completed the task accurately and 3 of them had small errors in their results. One of the participants missed some ticks at the export page, another used the correct shorthand on the wrong side, and the other used the wrong shorthand on the correct side. It is noteworthy that she used the same incorrect shorthand that she had used in Task3. 20% of the participants were confused about how to fill in the boxes in the corpus frequency, indicating a problem with the way corpus frequency might be searched. Only one of the participants stated that she was still confused with using the shorthand.

5.5 Summary

Of the requirements specified in the requirements gathering section, 4 out of 6 of them appeared to be met in this interface, namely the requirements concerning Export efficiency, Transcript efficiency, effectiveness, and satisfaction, Navigation bar consolidation, and Help specificity and usefulness. The last 2 requirements have been partially met as well. Users are able to conduct complex searches easily, confidently and efficiently, but the accuracy although improved from the old interface, is still unsatisfactory in the current prototype. The Layered Search appears to be satisfactorily usable, but its accessibility still needs a huge improvement.

It can be seen that there is still much to be improved in this interface, but the process of the user-centred redesign definitely helped to improve the usability of this interface. The next chapter will cover the main redesign phases of ElicitSpeech, the app associated with LaBB-CAT.

Chapter VI

ElicitSpeech Requirements Gathering

The next four chapters will cover the redesign process of ElicitSpeech. The process of requirements gathering for ElicitSpeech will be detailed in this chapter. The next two chapters will cover the three phases of the redesign, with each phase consisting of a design, prototyping, and evaluation component.

In this chapter the background and current interface of ElicitSpeech are detailed in Sections 6.1 and 6.2. The methodology for the usability testing performed is covered in Section 6.3, and the results obtained from the requirements gathering phase is described in Section 6.4. A summary of the requirements gathering process for LaBB-CAT is presented in Section 6.5.

6.1 Background of ElicitSpeech

ElicitSpeech is an application created to obtain specific speech recordings from participants with their consent. These data would then be sent to LaBB-CAT for researchers to access and analyse. The ultimate target audience of ElicitSpeech are people with dysarthria, but after discussion with the clients of the project, it was decided that the main goal for the first few phases of the project would be to make the application user-friendly for the average person. When the application is sufficiently user-friendly for the public, future redesign phases would only need to focus on ensuring that the prototype is also user-friendly for people with dysarthria. The main problem that might arise is that some people with dysarthria might also have motoric disabilities, which might affect their interaction with the application. With this kept in mind, the main button of the application is kept large and easy to press for users with disabilities.

6.2 *Current web interface of ElicitSpeech*

ElicitSpeech is an application that uses two different tasks, a sentence reading task and a map reading task, to obtain participants' speech. The application is simple, with the first few pages being dedicated for instructions, consent form, and demographics. The tasks included the display of several sentences on separate pages for 60 seconds, followed by a map reading task. As seen in Figures 6.1 to 6.4, the design of the interface is simple and plain. The fonts used are generally larger than those for LaBB-CAT, and the design is easier to learn and use.

In my first use of the platform, I found that the back button does not work on the browser when the web application is running, so I was unable to go back and re-read instructions on the previous page, or repeat a task that I made a mistake in. I also felt that the consent form is too lengthy and difficult to read 6.2. Also, as shown in 6.3, although it is clearly stated that the microphone access would be popping up above the page, it took me some time to realize what it meant. It might be better if a screenshot of the pop up was shown in the instructions page so that I would know what I am looking for. The small red timer showing a countdown of 60 seconds for each task was also a little intimidating at first, and it also took a while for me to understand what it was for 6.4. There was also no feedback or choice to listen to my recording when I completed a task, which was unsettling for me. On the other hand, I easily understood the tasks, and it took little to no effort to learn how to use the interface.

The current ElicitSpeech web platform interface is as shown in the screenshots below:

6.3 *Usability testing of ElicitSpeech's current interface*

As ElicitSpeech was built to record speech of dysarthric patients speaking a certain number of phrases, the ultimate target end users of ElicitSpeech are dysarthric patients. However, the project owners want to first find out the general usability of the current interface for the general population. Future research can then improve on the new interface and explore options with

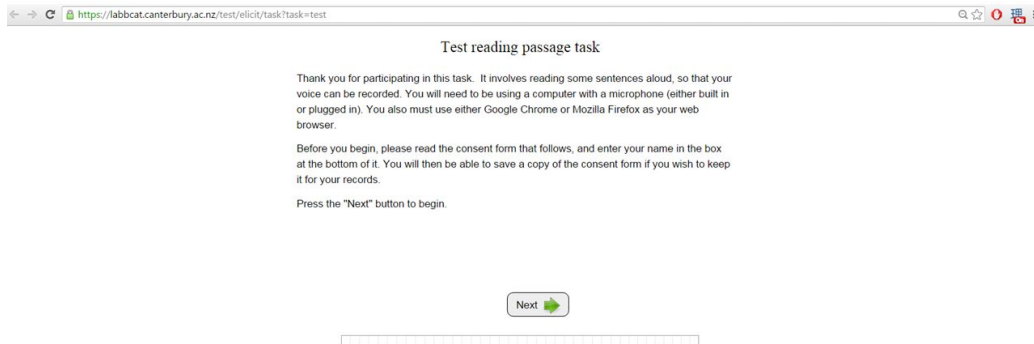


Figure 6.1: Home page of ElicitSpeech (web)

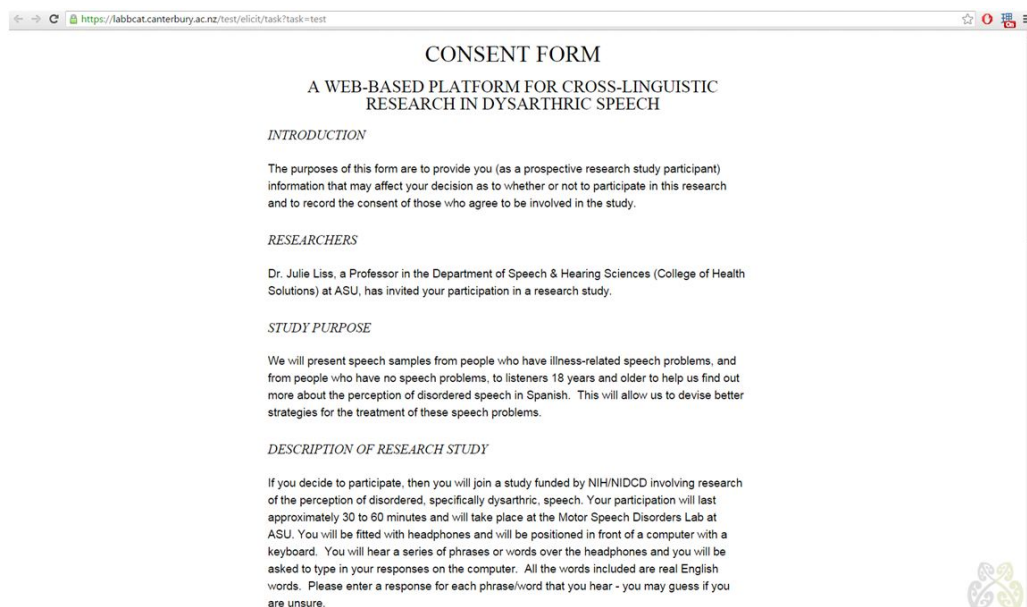


Figure 6.2: Consent form page of ElicitSpeech (web)

conducting a user-centred redesign with dysarthric patients.

6.3.1 Participants

Nine students at the University of Canterbury participated in the experiment. Of the participants, 77.8% were females and 22.2% were males. The mean age of the participants was 22.7 with a standard deviation of 4.90 in a range of 17 to 34 years old. The participants were recruited through posters placed in student advertisement locations all throughout the university.

Check Your Microphone

In the next step
you may be asked to enable access to your microphone,
which you do by clicking the **"Share Selected Device"** or
"Allow" button
on the box that pops up above the page.



Figure 6.3: Microphone setup of the ElicitSpeech (web)

Please read the following out loud:

Do you like amusement parks? Well, I
sure do.

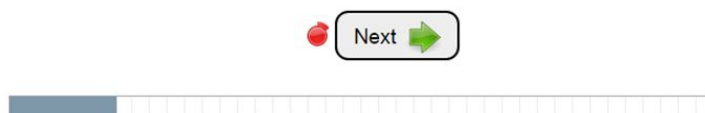


Figure 6.4: First task page of ElicitSpeech (web) (11 tasks in total)

6.3.2 Design and Apparatus

This study was conducted as requirements gathering through usability testing of the current interface. The dependent variables are the qualitative feedback obtained in the semi-structured interview. An iPad was provided with the ElicitSpeech app installed and opened.

6.3.3 Procedure

The study was conducted in a quiet and bright room in the University of Canterbury. Before the study began, participants were read an introduction paragraph ensuring them that the study is evaluating the interface and not them. It was specifically mentioned that there are not right or wrong answers to any of the questions asked in the study, and that they can feel free to speak their mind for the benefit of the product. They were briefed about being recorded for the study, and that they should think aloud throughout the entire session. They were then given the information sheet and consent form to sign.

Participants were given a semi-structured interview with open ended questions to ensure that they had little to no experience in design or coding. They were then instructed to complete the tasks on the ElicitSpeech application installed on the iPad provided. At completion, they were then given another semi-structured interview (Appendix C), which consisted of questions specifically targeted to find out problems from each of the different sections in the application (home page, consent form, mike enabling, sentence tasks, map tasks, last page), open-ended questions adapted from the System Usability Survey (SUS), as well as questions that allowed them to voice their opinions regarding what they liked and disliked about the interface. Upon completion of the study, they were awarded a \$5 university cafe voucher.

6.4 Results

Of the nine participants, all of them (100%) stated that the application was “simple”, “straightforward”, or “easy” to use. 77.8% of the total participants mentioned that the instructions were too lengthy or unclear. 66.7% of the

participants stated that the application was not fun, too “boring”, or needed colour. 33.3% of the participants wanted feedback from the application regarding what was recorded.

One of the participants did not know that they had to read out loud. One participant mentioned that she was not used to reading subway maps as New Zealand does not have any, which explained her confusion for that specific task. Another participant felt that she would require the support of a technical person to use the system as she was slightly confused when she initially began using the application. Three participants felt that there was inconsistency in the system, two of which was due to differences in fonts used within the system, and one was due to the huge amount of text shown in the consent form page. Two participants answered that they found the system cumbersome to use, however, both of them also mentioned in the open ended questions that the system was easy to use. This result could therefore be attributed to them being confused about what “cumbersome” meant.

A few suggestions were given by the participants for improving the application. One suggested to add pictures, and another suggested to include flashy and encouraging feedback to improve engagement.

6.5 Summary

The main findings from this user study were consistent with initial observations that the current interface is simple and easy to learn and use, but that there were too much unneeded instructions. The application lacked attractiveness, and users felt uncomfortable with the lack of feedback. Overall, most participants enjoyed the map task possibly because it was the only task that differed from the others, was more challenging, and had colour.

From these results, it was decided to improve the attractiveness of the interface, providing some form of feedback regarding mike recording, as well as removing the unnecessarily excessive amounts of text. Gamification elements could be added to improve attractiveness, which will be described in the next chapter.

Specific requirements for this interface are listed below:

1. Design needs to be visually appealing
2. User engagement needs to be improved

Chapter VII

ElicitSpeech Redesign Phase 1

This chapter will cover the first iteration of Design (Section 7.1), Prototype (Section 7.2), and Evaluation (Section 7.3). Results found from the evaluation are described in Section 7.4 and a summary of Phase 1 of the redesign for ElicitSpeech is covered in Section 7.5.

7.1 Design

As the current interface was sufficiently easy to use, the redesign focused on injecting gamification into the new interface to reduce boredom and increasing users' motivation to use and promote this application to others. Specifically, colours and images were added to the interface, and several game elements such as scoring and immediate feedback were included as well. It was also decided that the tasks used be more varied, including words, sentences, pictures and maps.

From the literature review of gamification, [37] states that designers should accept user motivational states as they are and push them towards the direction they would like to go and the place we want them to be. As participants would be given monetary incentives to use this application, it was decided to promote these incentives to improve their motivation. This would align user goals with our goals, which is them wanting to collect the monetary reward and completing all the tasks on the application.

Several sketches were made, and refined over time to be a single prototype in Balsamiq. These sketches can be seen in figures 7.1 to 7.2 below.

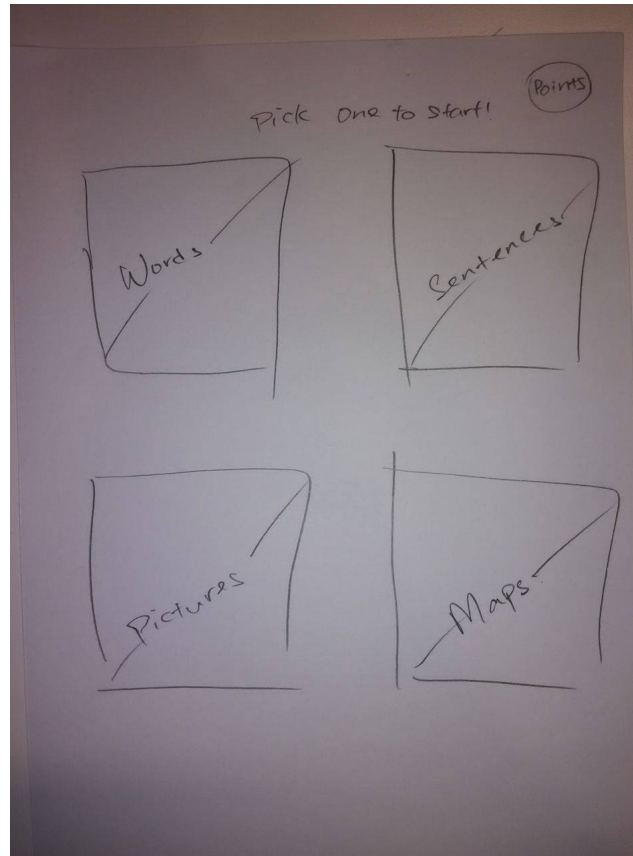


Figure 7.1: Sketch of Home page

7.2 Low-fidelity prototyping of *ElicitSpeech*

Low fidelity (Lo-Fi) prototypes were created in the early phases of the design process, following the requirements gathering user study. These prototypes are easy to create, and provides an easy way to obtain feedback about the design quickly. The initial low-fidelity prototypes were created using a combination of paper sketches and Balsamiq, a rapid wireframing tool.

Certain elements were also added to reduce boredom, such as the ability to choose the task they want to do 7.3, the rephrasing of instructions to include fun phrases, and the percentage system on the top right of the page to reflect the amount of money they have obtained for each task 7.4.

Finally, feedback was added to the prototype. Both feedback for the microphone recording 7.4 as well as motivational feedback for the end of the

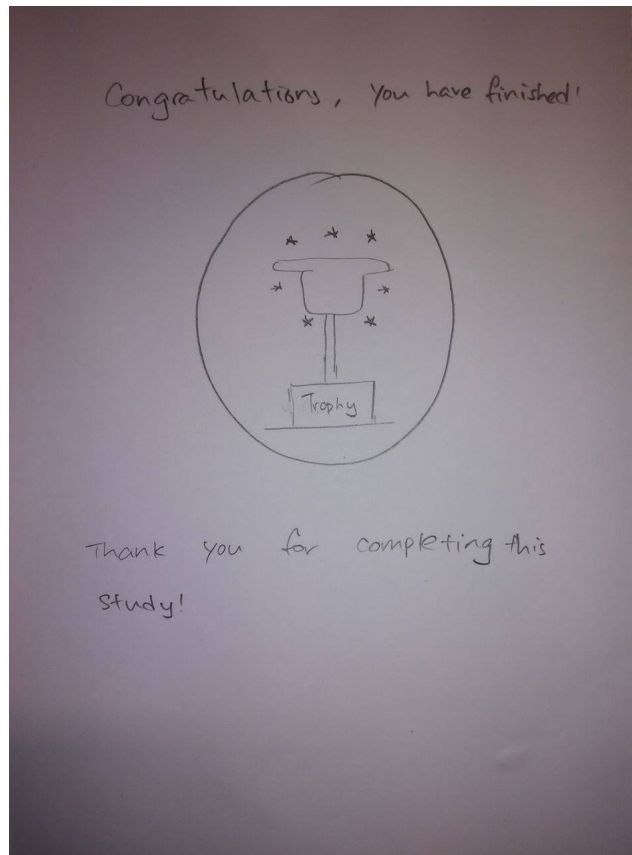


Figure 7.2: Sketch of Last page

application 7.5 was created and included in the prototype.

7.3 Usability testing of *ElicitSpeech's* low-fidelity prototype

7.3.1 Participants

Seven students at the University of Canterbury participated in the experiment. Of the participants, 85.7% were females and 14.3% were males. The mean age of the participants was 25.3 with a standard deviation of 3.68 in a range of 19 to 28 years old. The participants were recruited through posters placed in student advertisement locations all throughout the university.

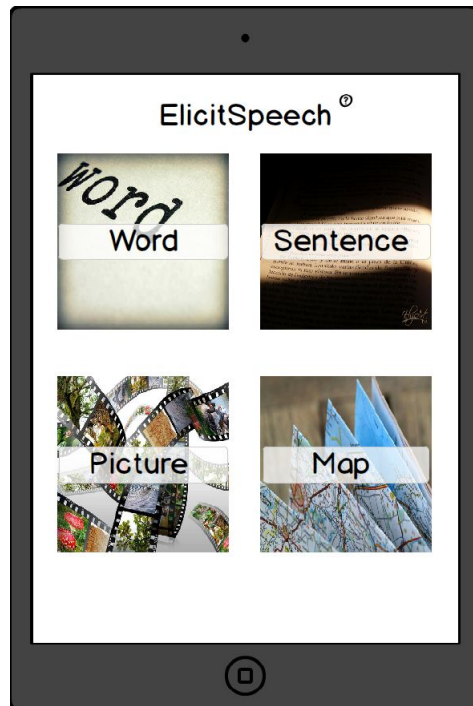


Figure 7.3: Home page used in the Lo-Fi Prototype of ElicitSpeech

7.3.2 Design and Apparatus

This study was conducted as requirements gathering through usability testing of the prototype. The prototype was opened in a PDF file on a Dell laptop with 15.6 inch screen.

7.3.3 Procedure

The study was conducted in a quiet and bright room in the University of Canterbury. Before the study began, participants were read an introduction paragraph ensuring them that the study is evaluating the interface and not them. It was specifically mentioned that there are not right or wrong answers to any of the questions asked in the study, and that they can feel free to speak their mind for the benefit of the product. They were briefed about being recorded for the study, and that they should think aloud throughout the entire session. They were then given the information sheet and consent form to sign.

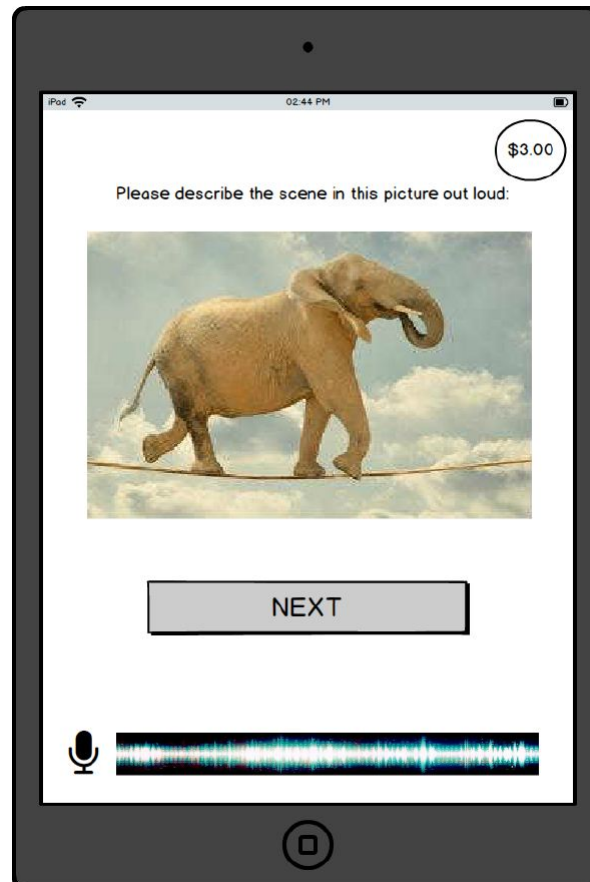


Figure 7.4: Example of picture task instructions used in the Lo-Fi Prototype of ElicitSpeech

Participants were given a semi-structured interview with open ended questions to ensure that they had little to no experience in design or coding. They were then instructed to complete the tasks on the ElicitSpeech application using the laptop provided. At completion, they were then given another semi-structured interview (Appendix C), which consisted of questions specifically targeted to find out problems from each of the different sections in the application (home page, consent form, mike enabling, sentence tasks, map tasks, last page), open-ended questions adapted from the System Usability Survey (SUS), as well as questions that allowed them to voice their opinions regarding what they liked and disliked about the interface. Upon completion of the study, they were awarded a \$5 university cafe voucher.

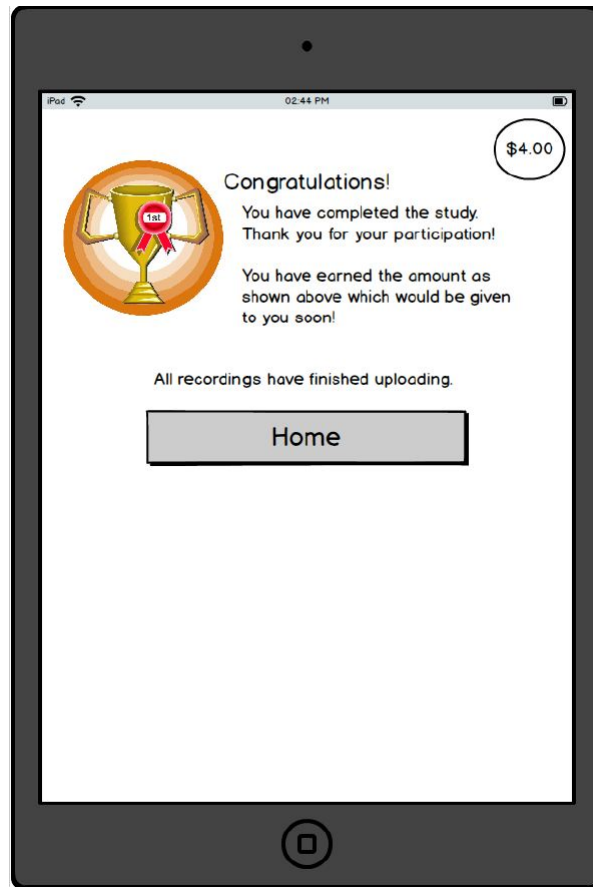


Figure 7.5: Last page used in the Lo-Fi Prototype of ElicitSpeech

7.4 Results

71.4% of the participants mentioned that the interface was simple, easy or straightforward. This is a drop from the old interface where 100% of the participants felt that the interface was easy to use. Moreover, there was no obvious display of confusion in the old interface, whereas most participants (85.7%) asked the experimenter for more instructions in for the new prototype at either the home page, last page, or about the instructions. Although only 2 participants (28.5%) mentioned that the prototype was boring or not fun, none of the participants made any mention of the interface being fun, interesting, or attractive in any way.

71.4% of the participants mentioned that they did not notice the money increasing for each task, and 28.5% of them also mentioned that they noticed

the money increasing, but did not pay much attention to it. This could either be due to the lack of animation or sounds to draw attention to the monetary rewards being given, or that the size of the icon was too small and out of sight.

7.5 Summary

Although the attractiveness of this prototype was appears to satisfactory, the prototype has declined in terms of ease of usability and learnability.

Due to this trade-off, it was decided that this prototype was not sufficiently good enough to replace the old one, and that a new low-fidelity prototype should be created that does not trade-off usability for attractiveness.

Chapter VIII

ElicitSpeech Redesign Phase 2

This chapter will cover the second iteration of Design (Section 8.1), Prototype (Section 8.2), and Evaluation (Section 8.3). Results found from the evaluation are described in Section 8.4 and a summary of Phase 2 of the redesign for ElicitSpeech is covered in Section 8.5.

8.1 Design

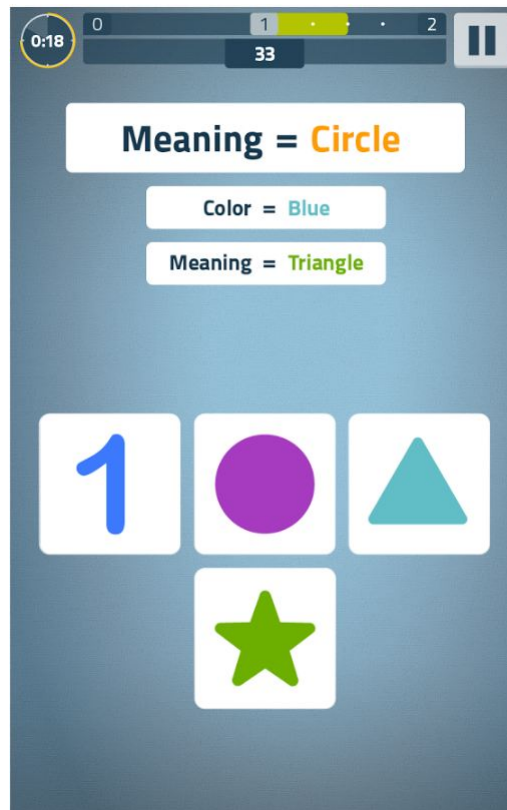


Figure 8.1: Sketch of Last page

According to Norman's 7 design principles [5, 3, 30], standardizing is a good way to ensure the success of systems. This sets me to research about similar applications available that have been successfully received by the public. One such product found was the NeuroNation application. NeuroNation is a web and mobile application that was designed to collect participant information through a the completion of a series of tasks, similar to what ElicitSpeech is built for. The main difference between NeuroNation and ElicitSpeech is that NeuroNation collects information about a user's current cognitive level and uses this data to create a personalized training plan for the brain, which to users would be a short-term reward. ElicitSpeech collects information about a user's current speech patterns and uses this data to aid in speech research, which to the actual target audience would be a long-term reward.

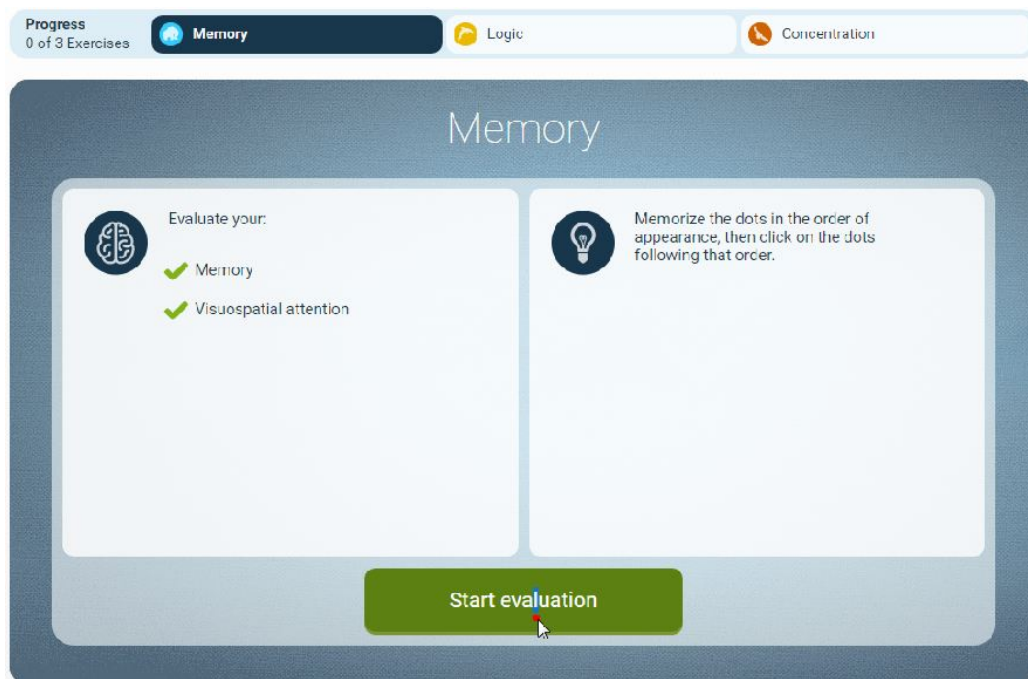


Figure 8.2: Instructions for tasks in NeuroNation Web Application

The NeuroNation mobile application (Figure 8.1) was rated 4.5 stars in the Google Play store, which shows its effective design. As the mobile application and the web application is very similar in terms of design, I decided

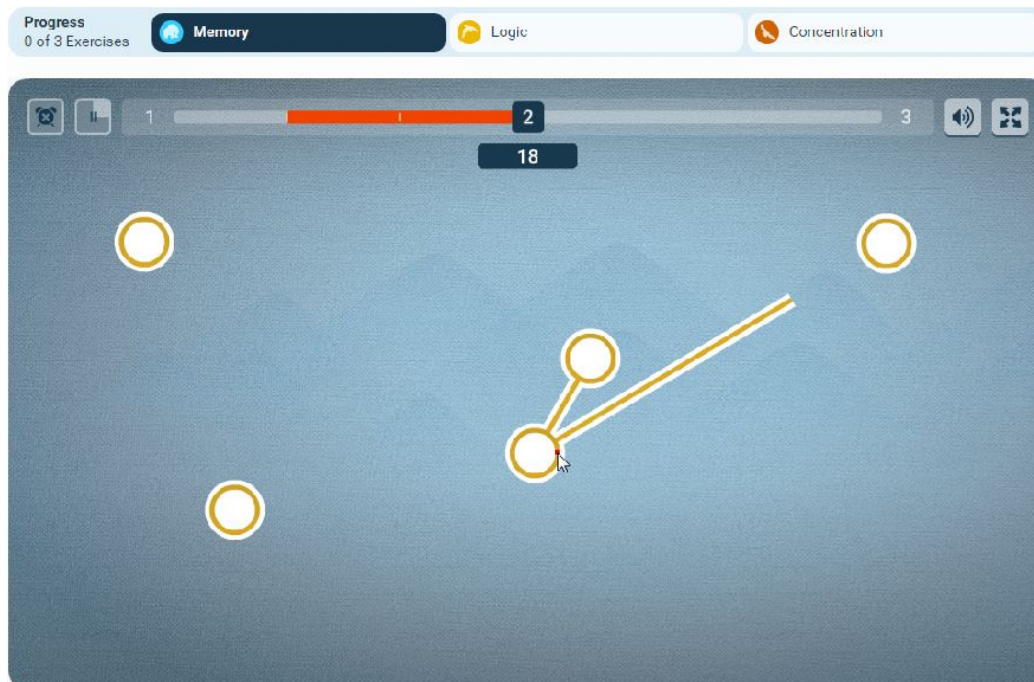


Figure 8.3: Example of task in NeuroNation Web Application

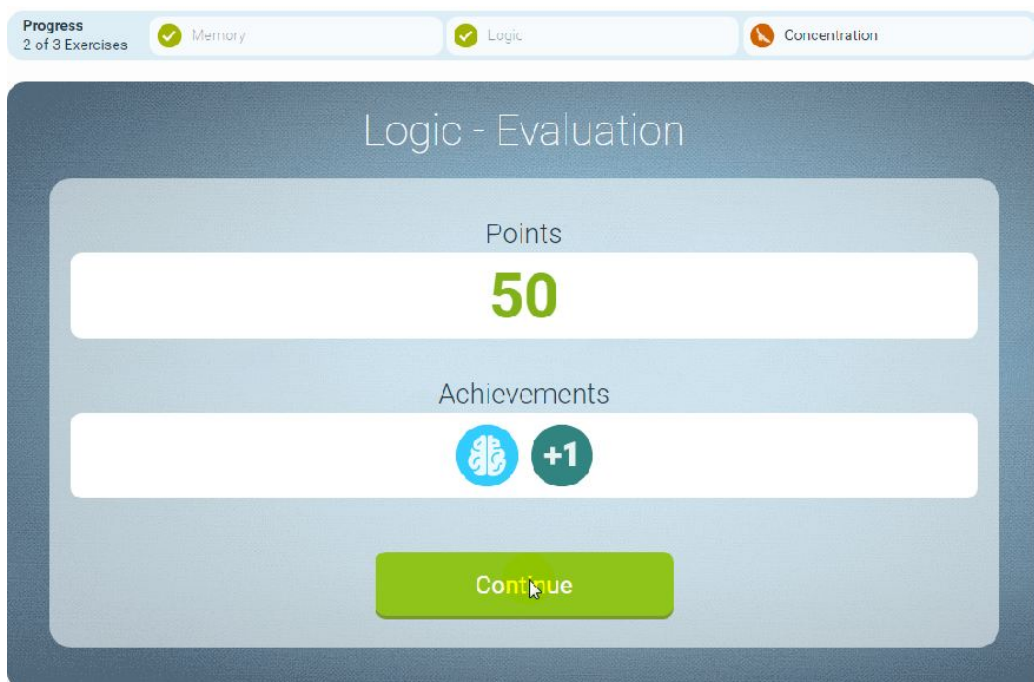


Figure 8.4: Feedback in NeuroNation Web Application

to create a paper prototype mockup of the web interface, replacing the tasks in NeuroNation with ElicitSpeech tasks.

Similar to the previous prototype, several sketches were created in the early phases of the design process. These sketches can be seen in figures 8.6 to 8.5 below.

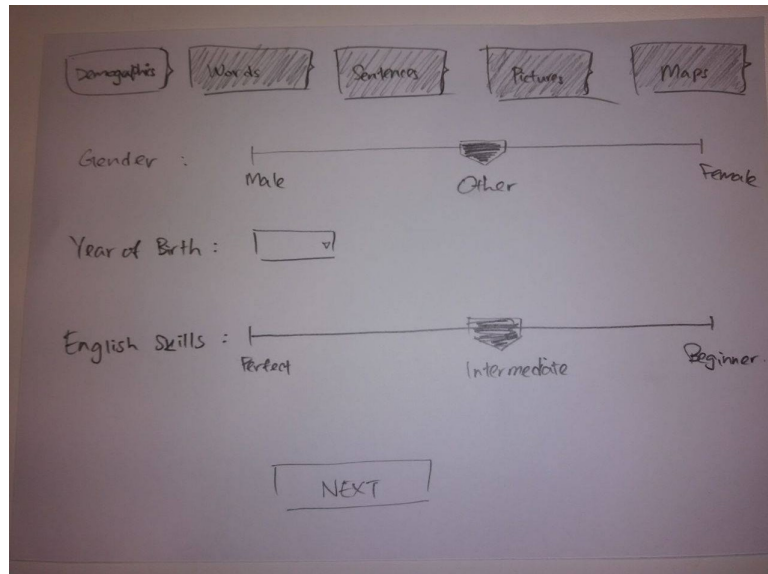


Figure 8.5: Sketch of Last page

8.2 Second low-fidelity prototyping of ElicitSpeech

This prototype was presented on paper instead of Balsamiq as some widgets required in this prototype was not interactive in Balsamiq. Images of the prototype can be seen in Figures 8.7, 8.8, and 8.9.

8.3 Usability testing of ElicitSpeech's second low-fidelity prototype

8.3.1 Participants

Seven students at the University of Canterbury participated in the experiment. Of the participants, 85.7% were females and 14.3% were males. The mean age of the participants was 25.3 with a standard deviation of 3.68 in a

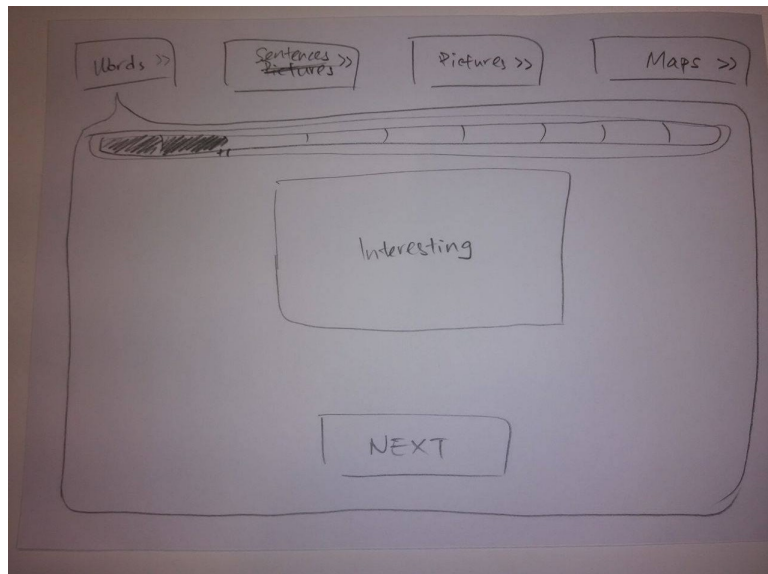


Figure 8.6: Sketch of Last page

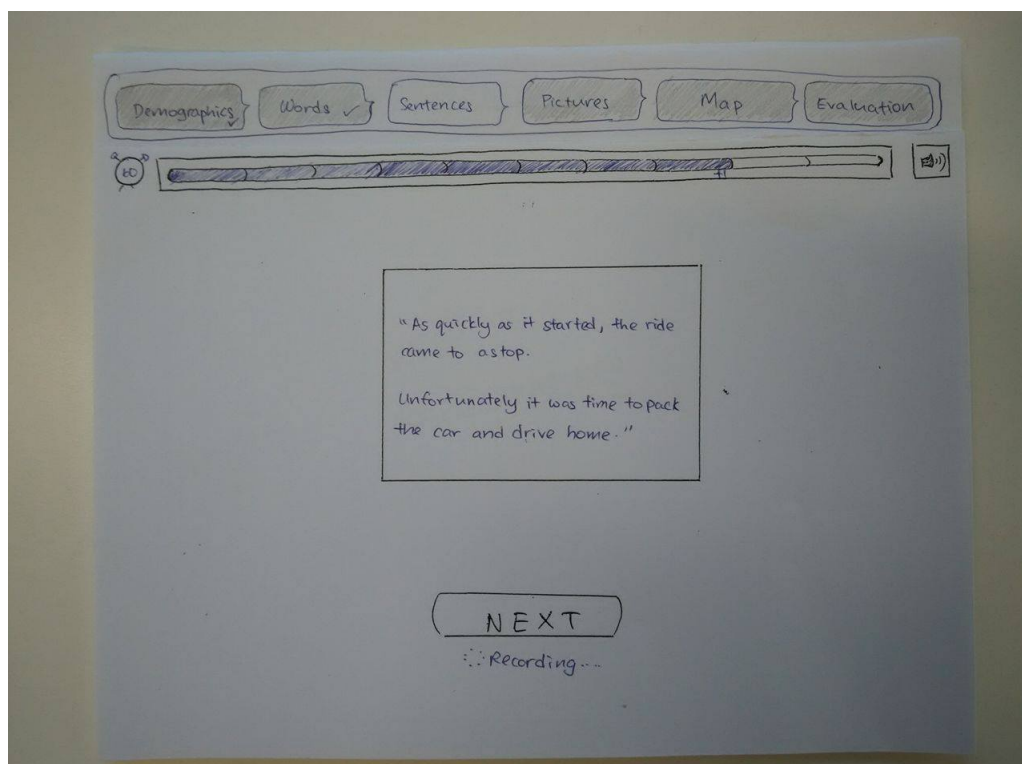


Figure 8.7: Example of sentence task used in the Lo-Fi Prototype of ElicitSpeech

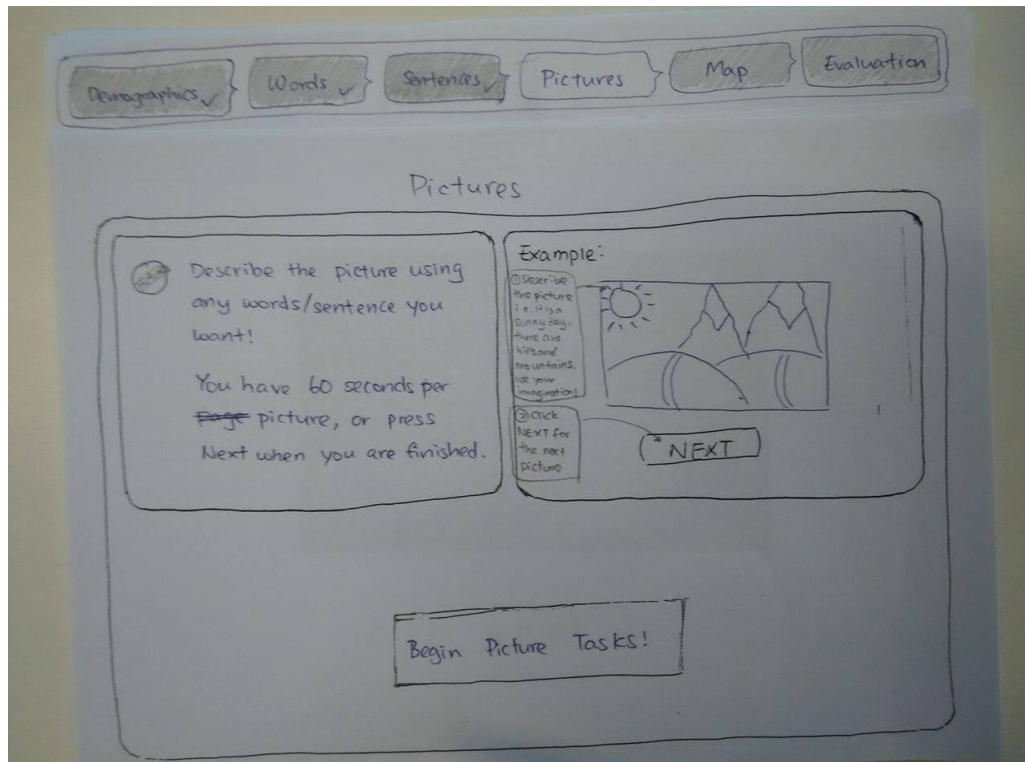


Figure 8.8: Example of picture task instructions used in the Lo-Fi Prototype of ElicitSpeech

range of 19 to 28 years old. The participants were recruited through posters placed in student advertisement locations all throughout the university, and had never used ElicitSpeech before.

8.3.2 Design and Apparatus

This study was conducted as requirements gathering through usability testing of the prototype. There was no independent variable, and the dependent variables are the qualitative feedback obtained in the semi-structured interview. The second low-fidelity paper prototype of ElicitSpeech was used in this study. The experimenter acted as the computer that the application was running on.

The User Experience Questionnaire (UEQ) [25] was used to investigate participants' impressions of the ElicitSpeech interface (Appendix D). It is used instead of the SUS survey due to its clear separation of the various

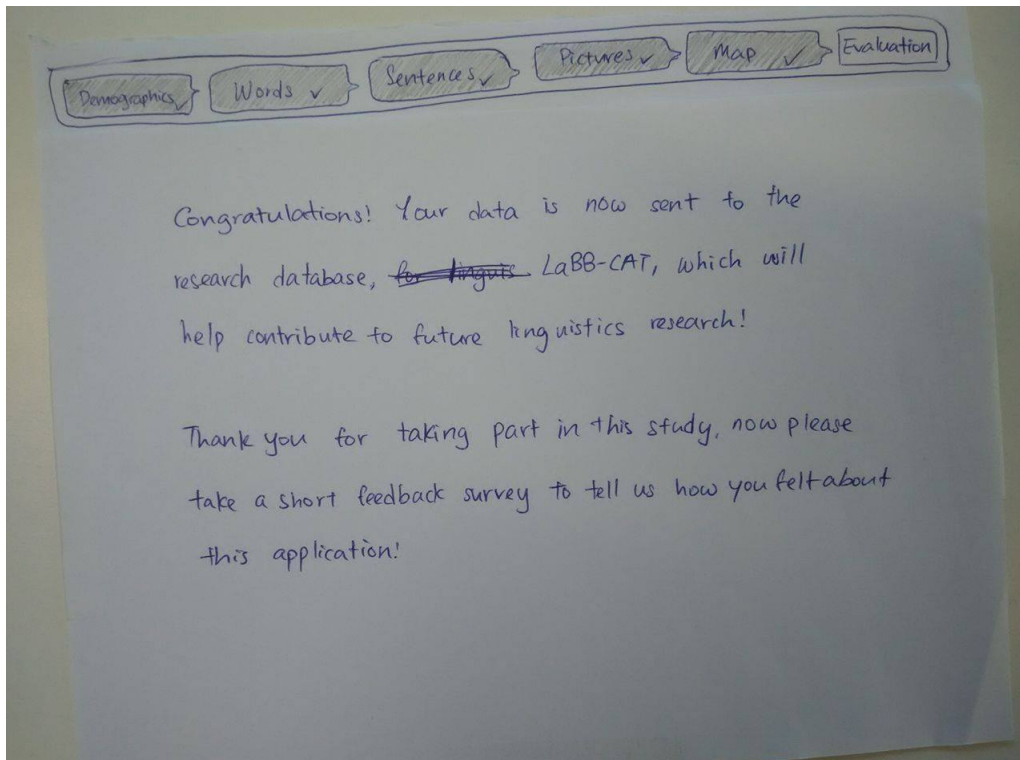


Figure 8.9: Last page used in the Lo-Fi Prototype of ElicitSpeech

aspects of user experience into 6 distinct scales.

It consists of 6 scales with 26 items, each having the form of a semantic differential as shown in Figure 8.10.

attractive o o o o o o o unattractive

Figure 8.10: Example of an item in the UEQ questionnaire

The 6 scales that can be measured with UEQ are Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty [36]. Attractiveness assesses the overall system impression of whether users like or dislike the system. Perspicuity measures the ease with which users familiarise themselves with the product, and the learning ease of the system. Efficiency determines the amount of effort that users require to solve the tasks. Dependability surveys the amount of control over the interactions users feel they have. Stimulation measures the level of excitement and motivation users feel when

using the system. Novelty quantifies the level of innovation and creativity of the system, and determines if the system is able to capture user interest.

The items in each scale are broken down in Figure 8.11 as shown.

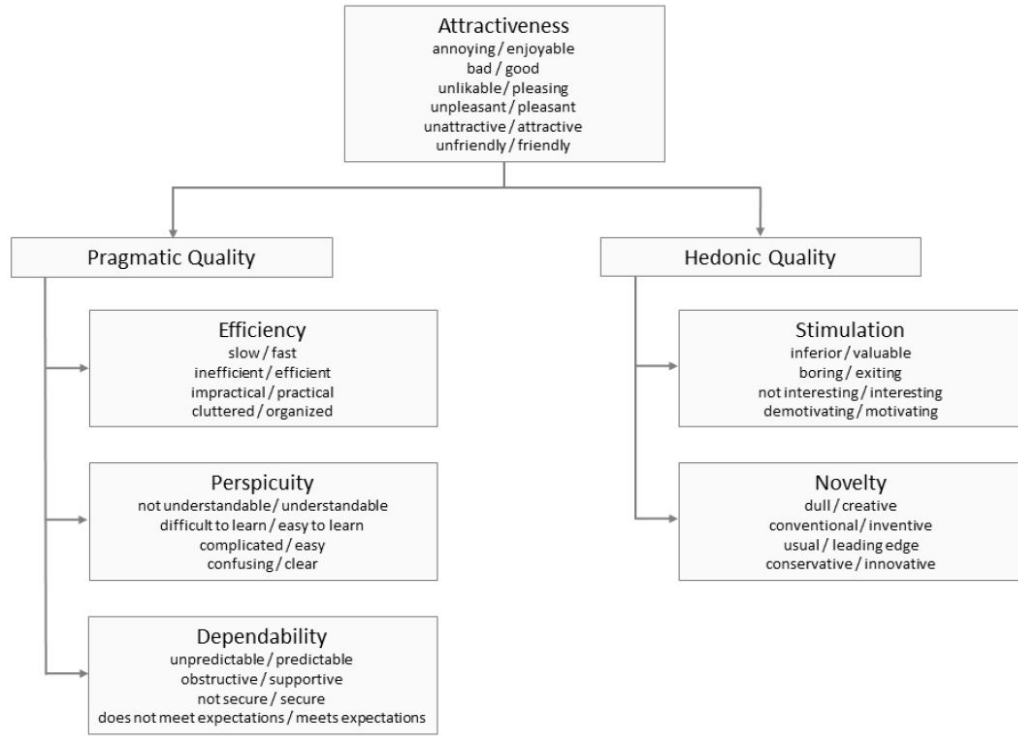


Figure 8.11: Structure and breakdown of items in the UEQ questionnaire [36]

8.3.3 Procedure

The study was conducted in a quiet and bright room in the University of Canterbury. Before the study began, participants were read an introduction paragraph ensuring them that the study is evaluating the interface and not them. It was specifically mentioned that there are not right or wrong answers to any of the questions asked in the study, and that they can feel free to speak their mind for the benefit of the product. They were briefed about being recorded for the study, and that they should think aloud throughout the entire session. They were then given the information sheet and consent form to sign.

Participants were given a semi-structured interview with open ended questions to ensure that they had little to no experience in design or coding. They were then instructed to complete the tasks on the ElicitSpeech application using the paper prototype. Similar to the LaBB-CAT paper prototype, participants were told that they could press anything they wish on the prototype. The experimenter would be standing at hand, and when the “Next” button was touched, the experimenter would remove the layers of paper one by one. At the completion of each task, a sound effect would be made that indicates an increase in points, and at the completion of an entire section, an extended version of the sound effect would be made.

At completion, they were then given another semi-structured interview (Appendix C), which consisted of questions specifically targeted to find out problems from each of the different sections in the application (home page, consent form, mike enabling, sentence tasks, map tasks, last page), open-ended questions adapted from the System Usability Survey (SUS), as well as questions that allowed them to voice their opinions regarding what they liked and disliked about the interface. Upon completion of the study, they were awarded a \$5 university cafe voucher.

8.4 Results

The scores for each of the UEQ scales are Attractiveness ($M=1.31$, $SD=1.01$), Perspicuity ($M=1.89$, $SD=1.10$), Efficiency ($M=1.71$, $SD=1.14$), Dependability ($M=1.18$, $SD=1.09$), Stimulation ($M=0.79$, $SD=0.82$), and Novelty ($M=0.43$, $SD=1.57$) respectively.

One of the participants (14.3%) mentioned that she found the product annoying as she generally did not like sounds that accompanied clicks and taps in applications. Two of the participants (28.6%) mentioned that the prototype had a lack of colours and looks common. Only one participant (14.3%) mentioned that the product was not very clear.

From my observations, I found that most of the participants (85.7%) were confused about the example shown on the Instructions page. Many of them did not realise that it was an example and tried pressing the example button, which yielded no action.

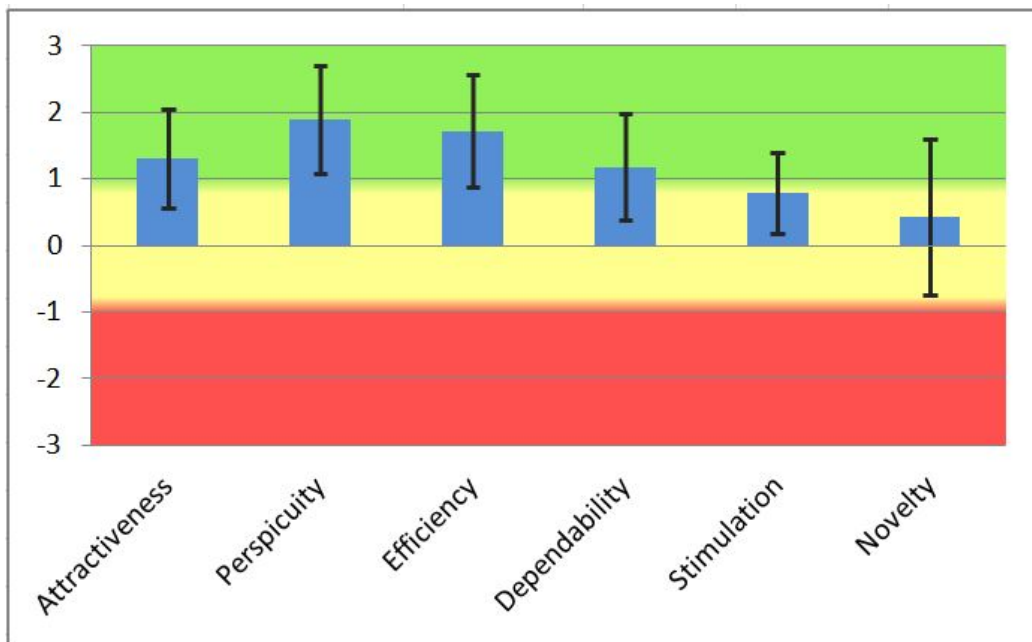


Figure 8.12: Graph of mean scores on the UEQ subscales

8.5 Summary

From the results, it was found that participants generally found the prototype to be likeable and easy to use. As the feedback obtained was generally positive, I decided that the design was sufficiently good enough to be built into a high-fidelity prototype. However, the application was still not perfect, and problems such as the confusing example on the instruction pages still need to be solved.

Chapter IX

ElicitSpeech Redesign Phase 3

This chapter will cover the third iteration of Design (Section 9.1), Prototype (Section 9.2), and Evaluation (Section 9.3). Similar to the previous chapter, results found from the evaluation are described in Section 9.4 and a summary of the redesign process for ElicitSpeech is covered in Section 9.5.

9.1 Design

As the Attractiveness, Perspicuity, and Efficiency of the paper prototype were all above average, I decided to proceed with the creation and evaluation of a High-Fidelity prototype.

Several changes were made to the prototype based on previous feedback. The example sections on the instructions page were not clear, which led to some confusion. These were therefore removed from the instructions page. Colour was also added to various aspects of the system as per the feedback obtained. Other parts of the system were retained similar to the paper prototype.

As the prototype was created in Axure, sound effects were unable to be added into the prototype. Moreover, as the prototype could not actually detect or record sounds from the users' microphone, a fake sound detection animation was added to the prototype to make users believe that the application was detecting sounds. This was created by displaying instructions asking users to attach their microphone to the computer, and that they should only click next when they are ready. They would then be shown a "Mic test bar" that would automatically change colour after 3 seconds similar to those in the sound test. To further ensure that users believed that the microphone was working, several users were asked to test out the interface, and none

of the users mentioned that they knew that the application was not really recording their voices.

9.2 High-fidelity prototyping of *ElicitSpeech*

The high fidelity prototype was designed using a combination of Photoshop and Axure as shown in Figures 9.1, 9.2, and 9.3.

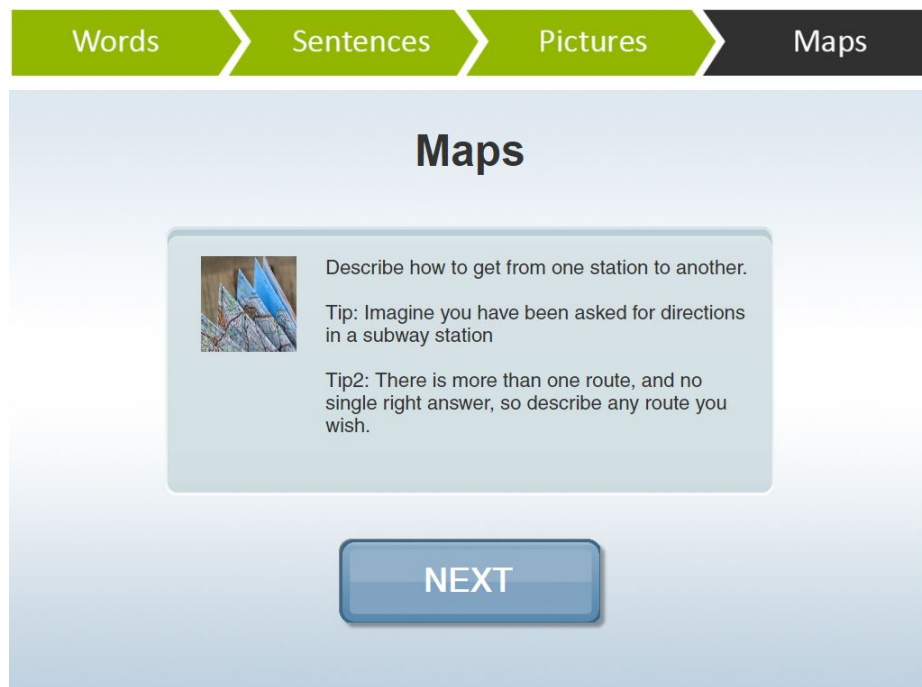


Figure 9.1: Home page in the Axure prototype of *ElicitSpeech*

9.3 Experimental evaluation of *ElicitSpeech*'s high-fidelity prototype

9.3.1 Setup

CrowdFlower was used to recruit participants for this study. Recruiting using such platforms eliminates the limitation that most researchers get stuck in, which is to recruit mostly university student volunteers as participants. These studies create the limitation of which the result would be limited to

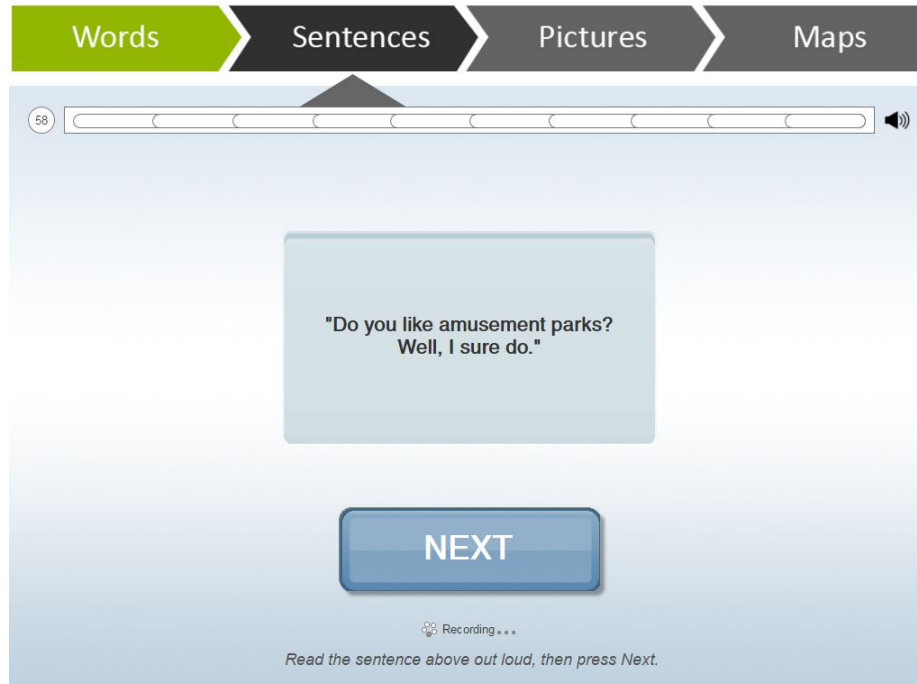


Figure 9.2: Example task in the Axure prototype of ElicitSpeech

the age range, literacy level, and location of the university. Crowdsourcing platforms source for volunteer workers from all over the world, with no limitations on age and literacy level. Results from crowdsourcing platforms were found in previous literature to be similar in quality to results that are obtained from traditional methods [14].

Participants were asked to provide demographics information, and to follow the instructions and complete the tasks of either the old or new interface of ElicitSpeech in a new popup window. They were instructed to obtain the participant ID information presented to them at the end of the application as proof of their completion, and to enter the ID into the CrowdFlower site to answer the UEQ survey (Appendix D). This study used a between-subjects design, and participants were randomly allocated to the different conditions, Old Interface or New Interface.

Each participant was only allowed to complete the study once, and had to spend at least 4 minutes on the entire study as that was the minimum amount of time required to complete the entire application and survey.

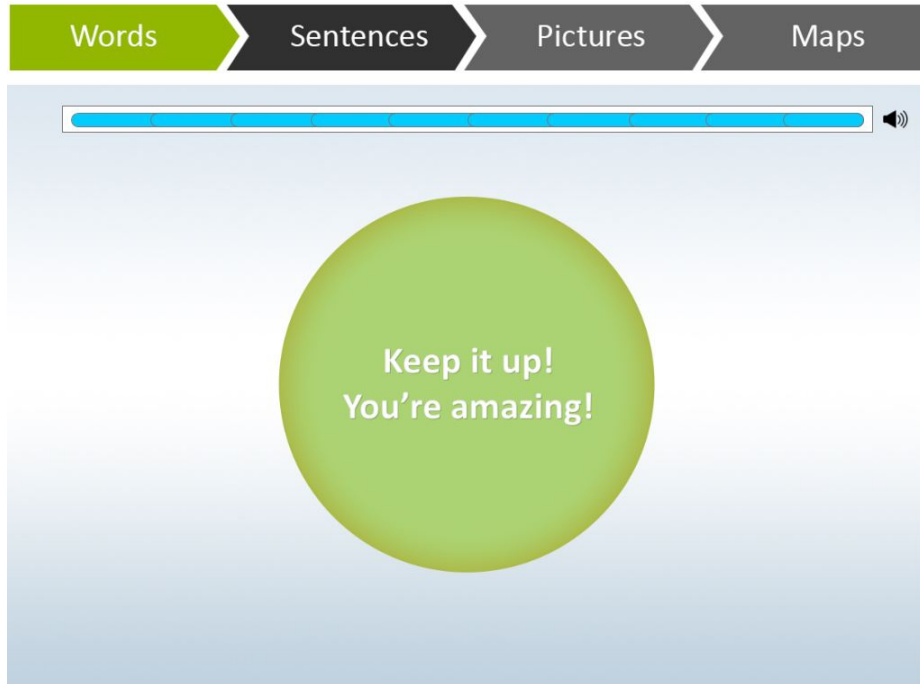


Figure 9.3: Motivation in the Axure prototype of ElicitSpeech

Participants that answered more quickly than this had their responses removed from the study. As some participants logged into different accounts to complete this study multiple times, the IP address of the participants were also logged, and responses with shared IP addresses were removed as well. This ensured the quality of the responses obtained.

As the resulting data obtained was fewer than 20 per condition, another similar study was conducted on Facebook. Participants were obtained through the sharing of the link on the experimenter's Facebook wall. Similar to participants recruited through the CrowdFlower site, participants were asked to follow the instructions and complete the tasks of either the old or new interface of ElicitSpeech in a new popup window. They were asked to return to the Qualtrics site to answer the UEQ survey (Appendix D). This study also used a between-subjects design, and participants were also randomly allocated to the two different conditions.

As the study was conducted online, a heuristic used by UEQ to ensure quality was to check the difference between the best and worst evaluation of

an item in a scale. If there was a big difference of more than three scales having inconsistent data, the data would be considered suspicious data and should be removed. Following this heuristic, nine responses were removed from our data set.

9.3.2 Participants

60 participants were recruited online using CrowdFlower and 28 participants were recruited online through Facebook. The participants from CrowdFlower were recruited to participate willingly using a variety of channels on CrowdFlower, and were paid USD\$1.00 for the completion of the entire study. Participants were allowed to give a contributor satisfaction rating at the conclusion of the study within the Crowdfower system. The mean contributor satisfaction rating obtained for this study was 4.3 out of 5, indicating that participants felt that the pay and workload was reasonable. The participants from Facebook were recruited to participate willingly through the sharing of the link on the Experimenter’s Facebook Wall. Participants obtained through Facebook were not paid for the completion of the study.

Of the 88 responses obtained, 16 had duplicate IP addresses with previous responses and were therefore removed. Of the 72 remaining responses, 5 had a completion time of less than 4 minutes and were also removed from the data analysis.

Of the 67 remaining participants, 67.2% were males and 32.8% were females. The mean age of the participants was 31.3 with a standard deviation of 9.49 in a range of 20 to 60 years old. Participants from CrowdFlower spent an average of 9 minutes on the study, while the data of the time taken for the study was not obtained for participants recruited from Facebook.

9.4 Results and Discussion

9.4.1 Comparison of data obtained from CrowdFlower and FaceBook

To ensure that the data from obtained from participants recruited from CrowdFlower and FaceBook were not significantly different, the means of the samples were compared. The Shapiro-Wilk normality test (Appendix F,

Figure F.1) found that most of the data were significantly normal, and those that were not significantly normal are shown in bold. The visual inspection of the Q-Q plots (Figure 9.4) for these data found that they were close enough to normal to be assumed normal.

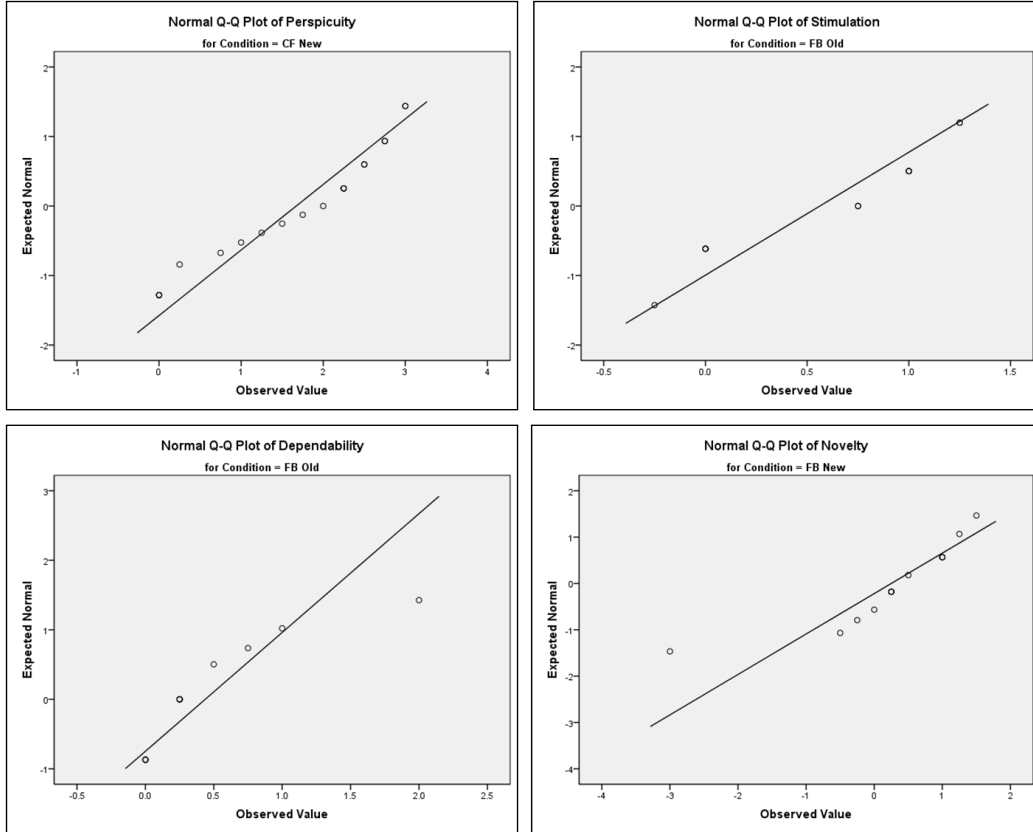


Figure 9.4: Q-Q plots of CrowdFlower and FaceBook data. CrowdFlower New Interface for Perspicuity (Top left), FaceBook Old Interface for Stimulation (Top right), Facebook Old Interface for Dependability (Bottom left), Facebook New Interface for Novelty (Bottom right).

As the data was found to be normal, a T-Test for independent samples was conducted between CrowdFlower participants ($N=19$) and Facebook participants ($N=13$) for the ratings for the New Interface (Appendix F, Figure F.2, Figure F.3). For the New Interface condition, there was no significant difference in Attractiveness ($t(30)=.95$, $p=.35$) rated by CrowdFlower participants ($M=1.77$, $SD=1.09$) and Facebook participants ($M=1.37$, $SD=1.29$). There was also no significant difference in Perspicuity ($t(30)=.18$, $p=.86$) rated by

CrowdFlower participants ($M=1.67$, $SD=1.06$) and Facebook participants ($M=1.60$, $SD=1.26$). There was also no significant difference in Efficiency ($t(30)=-.22$, $p=.83$) rated by CrowdFlower participants ($M=1.51$, $SD=1.02$) and Facebook participants ($M=1.60$, $SD=1.12$). There was also no significant difference in Dependability ($t(30)=.42$, $p=.68$) rated by CrowdFlower participants ($M=1.17$, $SD=1.03$) and Facebook participants ($M=1.00$, $SD=1.27$). There was also no significant difference in Stimulation ($t(30)=1.19$, $p=.24$) rated by CrowdFlower participants ($M=1.55$, $SD=1.17$) and Facebook participants ($M=1.00$, $SD=1.45$). There was also no significant difference in Novelty ($t(30)=1.61$, $p=.12$) rated by CrowdFlower participants ($M=.99$, $SD=1.35$) and Facebook participants ($M=.25$, $SD=1.14$).

A T-Test for independent samples was also conducted between CrowdFlower participants ($N=14$) and Facebook participants ($N=12$) for the ratings for the Old Interface condition (Appendix F, Figure F.4, Figure F.5). For the Old Interface condition, there was no significant difference in Attractiveness ($t(24)=1.47$, $p=.16$) rated by CrowdFlower participants ($M=1.27$, $SD=.97$) and Facebook participants ($M=.78$, $SD=.71$). There was also no significant difference in Perspicuity ($t(24)=-.20$, $p=.85$) rated by CrowdFlower participants ($M=1.13$, $SD=1.13$) and Facebook participants ($M=1.20$, $SD=1.00$). There was also no significant difference in Efficiency ($t(24)=.33$, $p=.75$) rated by CrowdFlower participants ($M=.96$, $SD=1.00$) and Facebook participants ($M=.85$, $SD=.63$). There was also no significant difference in Dependability ($t(24)=1.69$, $p=.10$) rated by CrowdFlower participants ($M=.89$, $SD=.76$) and Facebook participants ($M=.44$, $SD=.59$). There was also no significant difference in Stimulation ($t(24)=1.36$, $p=.19$) rated by CrowdFlower participants ($M=1.02$, $SD=1.03$) and Facebook participants ($M=.56$, $SD=.57$). There was also no significant difference in Novelty ($t(24)=.87$, $p=.06$) rated by CrowdFlower participants ($M=.96$, $SD=.85$) and Facebook participants ($M=.31$, $SD=.82$).

Since there was no significant differences in the data obtained from CrowdFlower participants and Facebook participants in both the new and old interface, the data were considered to be sufficiently similar to be combined for data analysis.

9.4.2 Comparison of Old interface and New prototype

To find out whether there was a significant difference in the ratings for Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty between the old and new interface, the means of the samples were compared. The Shapiro-Wilk normality test (Appendix F, Figure F.6) found that most of the data were significantly normal, and those that were not significantly normal are shown in bold. The visual inspection of the Q-Q plots (Figure 9.5) for these data found that they were close enough to normal to be assumed normal.

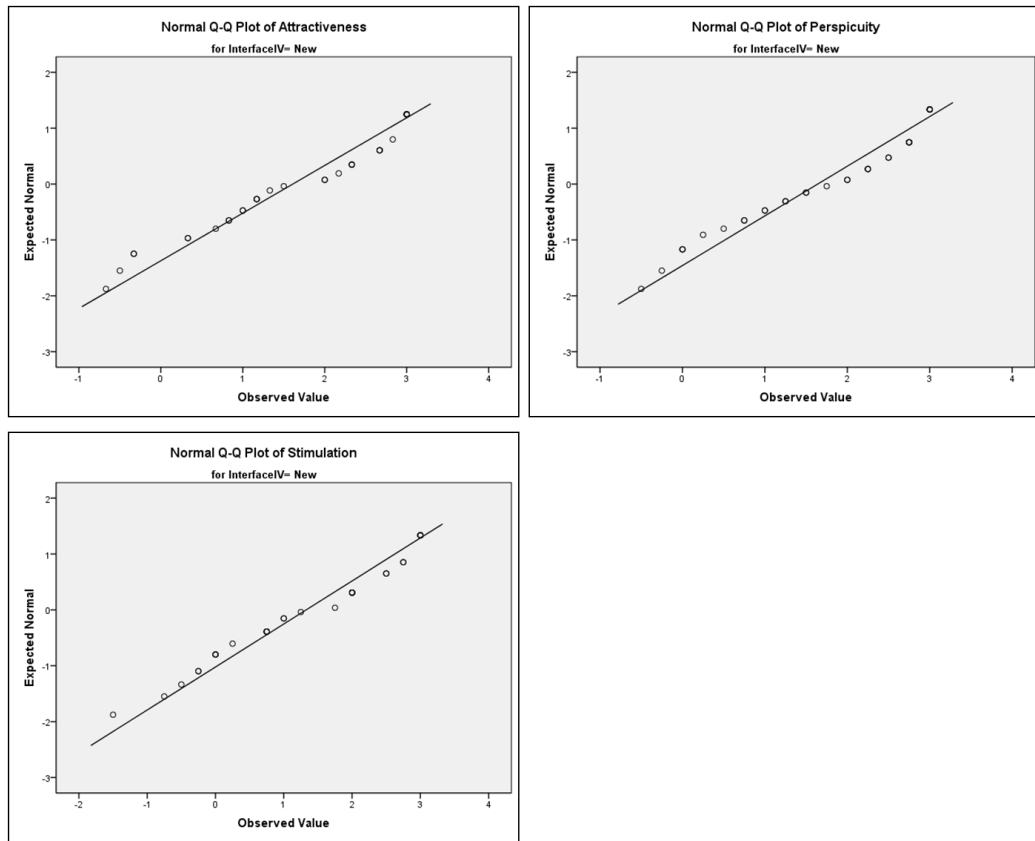


Figure 9.5: Q-Q plots of Old and New interface. New Interface for Attractiveness (Top left), New Interface for Perspicuity (Top right), New Interface for Stimulation (Bottom left),

As the data was found to be normal, a T-Test for independent samples was conducted between UEQ ratings for the Old Interface (N=26) and UEQ

ratings for the New Interface (N=32) (Appendix F, Figure F.7, Figure F.8).

Scale	New Interface						Old Interface					
	Mean	STD	N	Confidence	Confidence Interval		Mean	STD	N	Confidence	Confidence Interval	
Attractiveness	1.61	1.17	32	0.41	1.20	2.02	1.04	0.88	26	0.34	0.71	1.38
Perspicuity	1.64	1.13	32	0.39	1.25	2.03	1.16	1.05	26	0.40	0.76	1.57
Efficiency	1.55	1.05	32	0.36	1.18	1.91	0.91	0.84	26	0.32	0.59	1.23
Dependability	1.10	1.12	32	0.39	0.71	1.49	0.68	0.71	26	0.27	0.41	0.96
Stimulation	1.33	1.30	32	0.45	0.88	1.78	0.81	0.86	26	0.33	0.48	1.14
Novelty	0.69	1.30	32	0.45	0.24	1.14	0.66	0.88	26	0.34	0.32	1.00

Figure 9.6: Table of UEQ scores of old and new interface conditions

There was a statistically significant difference in the Attractiveness scores for the old interface (M=1.05, SD=.88) and new interface (M=1.61, SD=1.17) conditions; $t(55.7)=-2.04$, $p=.04$ (Figures 9.7 and 9.6). There was also a statistically significant difference in the Efficiency scores for the old interface (M=.91, SD=.84) and new interface (M=1.55, SD=1.05) conditions; $t(56)=-2.50$, $p=.02$ (Figures 9.8 and 9.6).

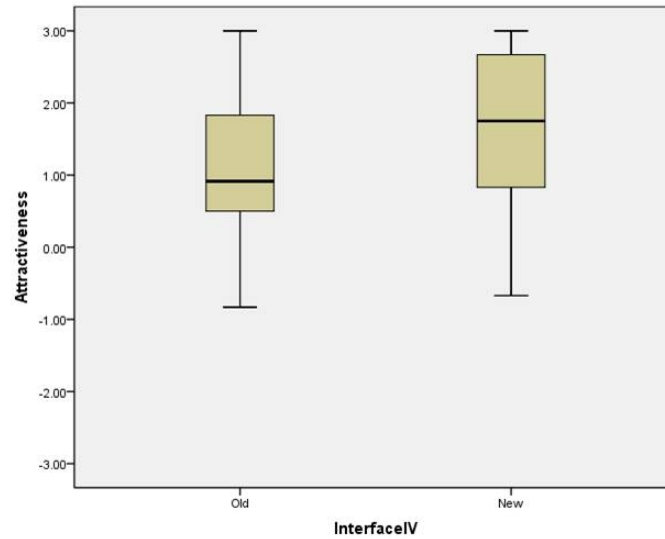


Figure 9.7: Box plot of comparison between old interface and new prototype in Attractiveness subscale

There was an established trend towards a significant difference in the Perspicuity scores for the old interface (M=1.16, SD=1.05) and new interface (M=1.64, SD=1.13) conditions; $t(56)=-1.65$, $p=.10$ (Figures 9.9 and 9.6). There was also an established trend towards a significant difference in the

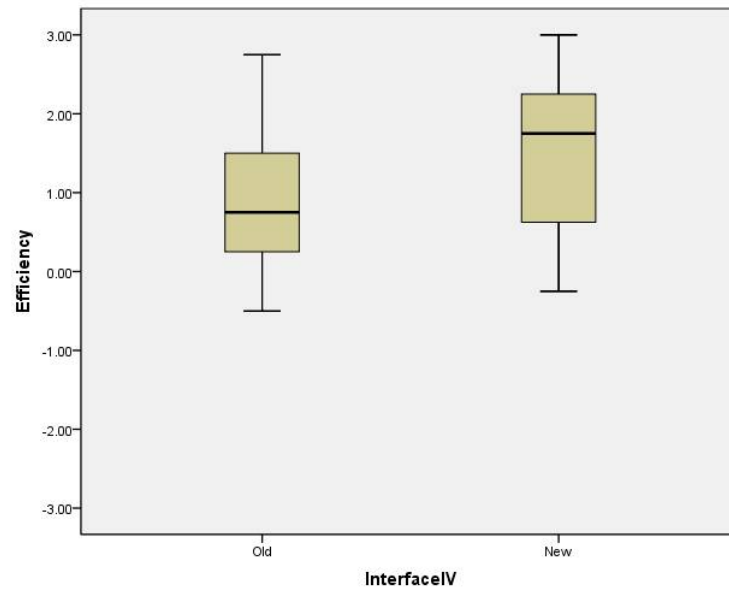


Figure 9.8: Box plot of comparison between old interface and new prototype in Efficiency subscale

Stimulation scores for the old interface ($M=.808$, $SD=.864$) and new interface ($M=1.33$, $SD=1.30$) conditions; $t(54.1)=-1.82$, $p=.07$ (Figures 9.10 and 9.6).

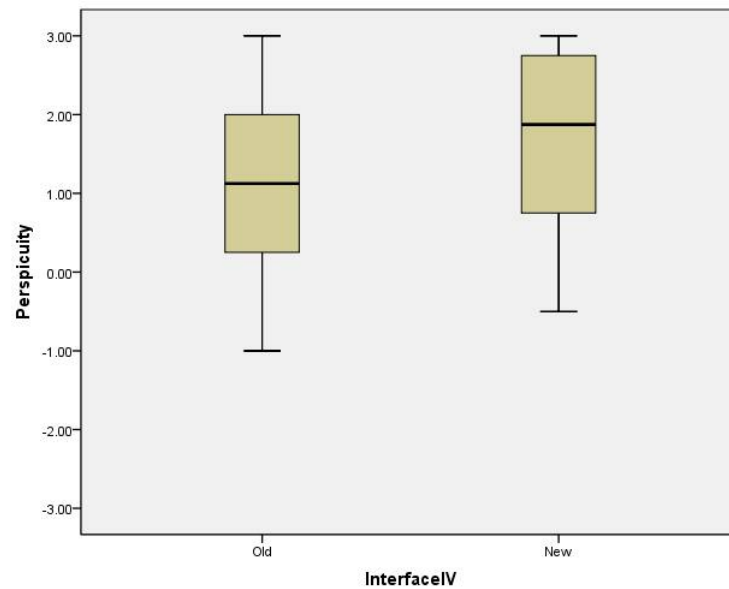


Figure 9.9: Box plot of comparison between old interface and new prototype in Perspicuity subscale

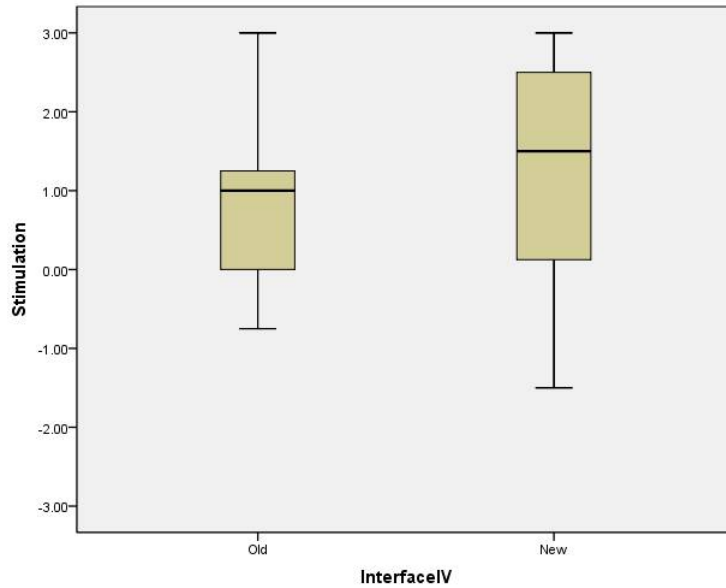


Figure 9.10: Box plot of comparison between old interface and new prototype in Stimulation subscale

There was an apparent trend towards significant difference in the Dependability scores for the old interface ($M=.68$, $SD=.86$) and new interface ($M=1.10$, $SD=1.12$) conditions; $t(53.1)=-1.73$, $p=.09$ (Figures 9.11 and 9.6).

There was also no significant difference in the Novelty scores for the old interface ($M=.66$, $SD=.88$) and new interface ($M=.69$, $SD=1.30$) conditions; $t(56)=-.08$, $p=.94$ (Figures 9.12 and 9.6).

Comparison of the mean scores of the new and old interface in the UEQ scales can be seen in Figure 9.13. As seen clearly in the graph, there was clearly a higher mean score for the new interface as compared to the old interface for all the UEQ scales except for Novelty.

There was a significant difference in the UEQ scores for Attractiveness and Efficiency between the old and new interface conditions. This could mean that the visual elements incorporated into the new interface was deemed as visually appealing by users. Moreover, the significant difference in Efficiency indicates that the gamification elements incorporated might have created flow, which tricked participants into believing less time has passed. This is further supported as there was no difference in the mean time taken to complete the study between the old interface (10 minutes) and new interface

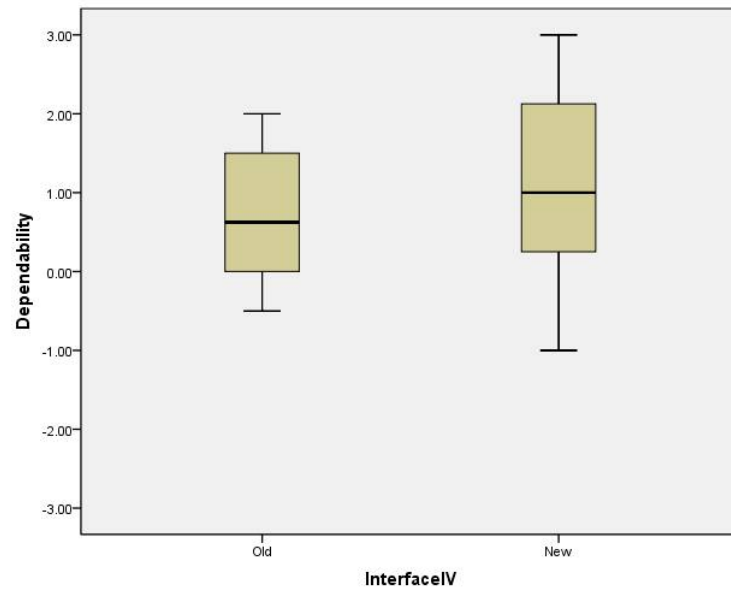


Figure 9.11: Box plot of comparison between old interface and new prototype in Dependability subscale

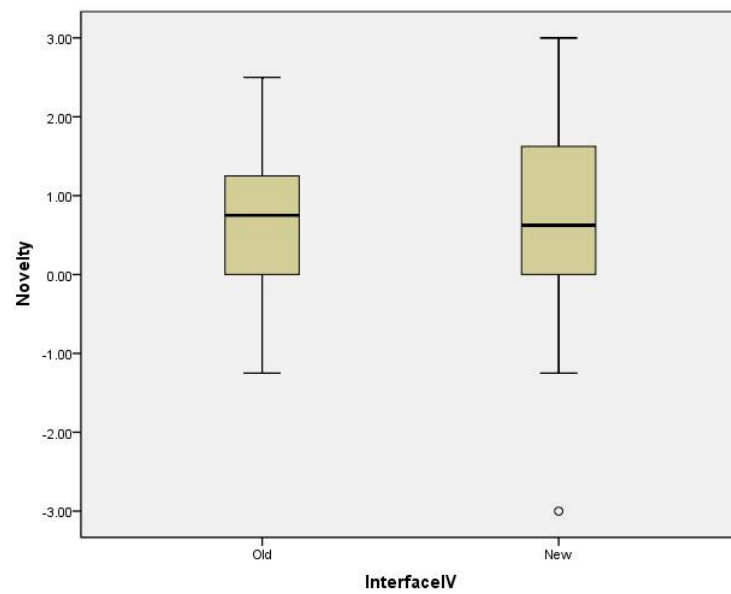


Figure 9.12: Box plot of comparison between old interface and new prototype in Novelty subscale

(9 minutes).

There was no significant difference in the UEQ scores for Perspicuity, De-

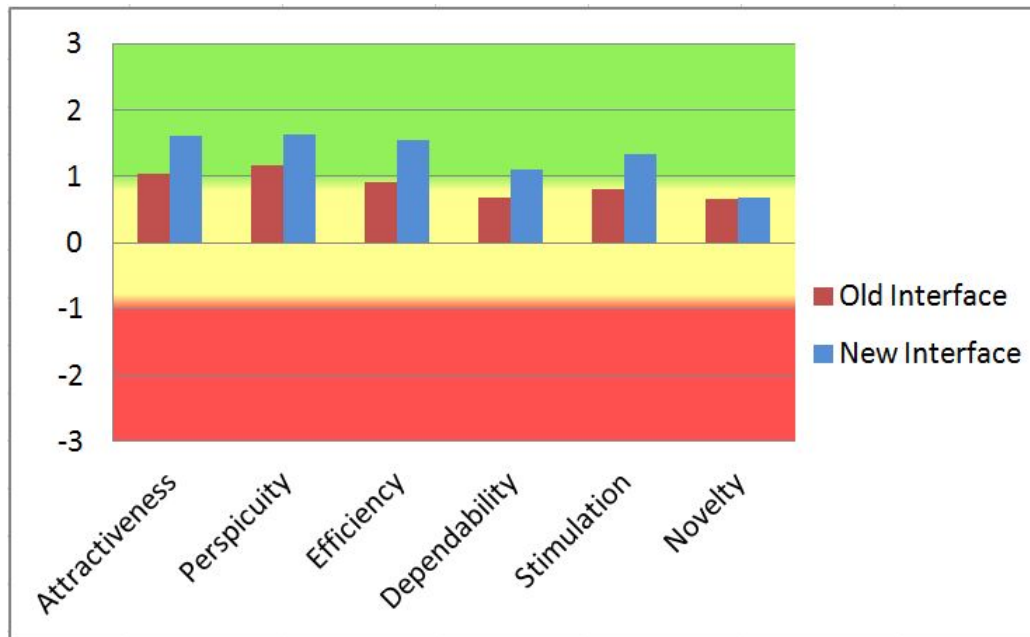


Figure 9.13: Graph of UEQ scores of old and new interface conditions

pendability, Stimulation, and Novelty. As mentioned in Chapter 6, the old interface was already sufficiently clear and easy. As there was no decrease in the Perspicuity of the interface with most of the instructions shortened, this shows that the instructions provided was sufficient, and could possibly be further reduced in the future. Moreover, as the prototype created did not actually detect and record sounds from the microphone, the low dependability of the prototype could be attributed to participants who were confused about whether the application was really recording their voice. The lack of significant difference in Stimulation indicates an area for future improvements, but it could be possible that the current amount of stimulation is sufficient for people with dysarthria to prevent seizures or over-stimulation, which could result in negative side effects. A possibility for the lack of difference in novelty could be due to the application being considered “Novel” only by nature of the voice recording task that is present in both the old and new interface applications, which is to record user’s voice by speaking out loud.

9.4.3 Benchmark

The UEQ survey comes with a benchmark data set containing data from 163 studies of various products. This allows for the comparison of the evaluated system with the benchmark to show the relative quality compared to other products in the market.

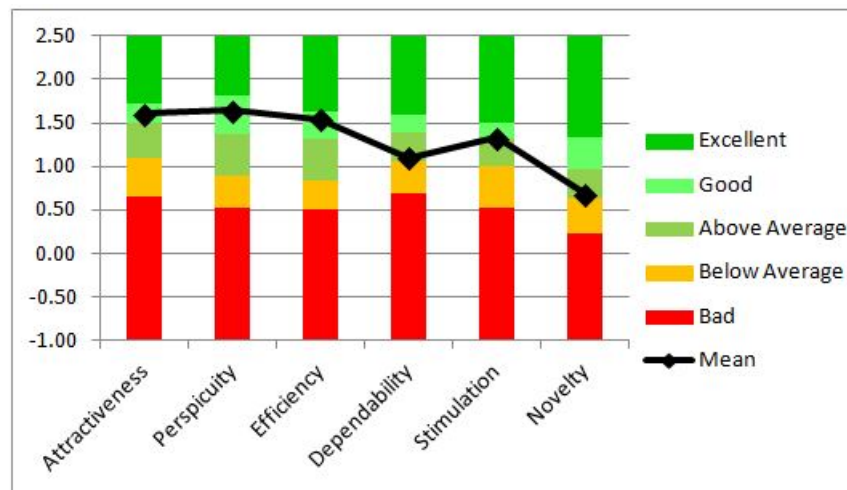


Figure 9.14: UEQ Benchmark of new interface

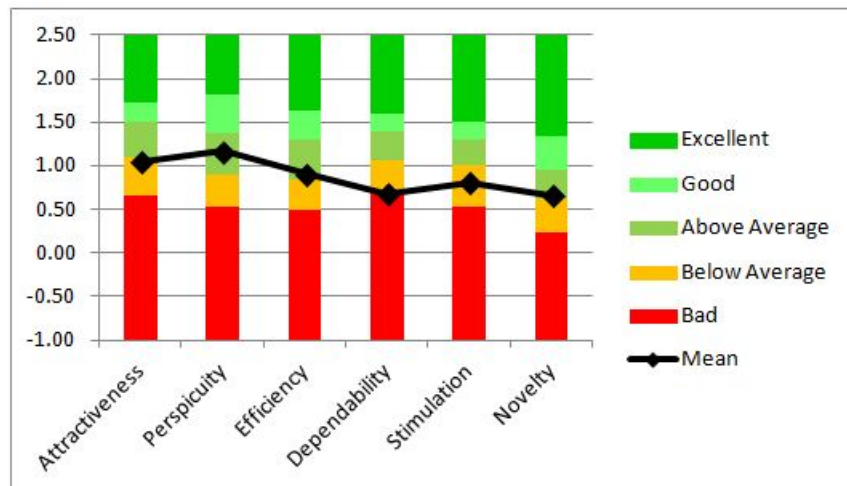


Figure 9.15: UEQ Benchmark of old interface

As shown in Figures 9.14 and 9.16, it was found that in the benchmark of the new prototype against the other products rated using the UEQ survey,

the new prototype was ranked Good for Attractiveness, Perspicuity, and Efficiency, and Above Average for Dependability, Stimulation, and Novelty. This reflects that the product yields a sufficiently good user experience, and therefore this redesign of ElicitSpeech can be considered successful.

Table of benchmark of New Interface

Scale	Mean	Comparisson to bench	Interpretation
Attractiveness	1.61	Good	10% of results better, 75% of results worse
Perspiciuity	1.64	Good	10% of results better, 75% of results worse
Efficiency	1.55	Good	10% of results better, 75% of results worse
Dependability	1.10	Above Average	25% of results better, 50% of results worse
Stimulation	1.33	Above Average	10% of results better, 75% of results worse
Novelty	0.69	Above Average	25% of results better, 50% of results worse

Figure 9.16: Table of UEQ Benchmark of new interface

Table of benchmark of Old interface

Scale	Mean	Comparisson to benchmark	Interpretation
Attractiveness	1.04	Below average	50% of results better, 25% of results worse
Perspiciuity	1.16	Above Average	25% of results better, 50% of results worse
Efficiency	0.91	Above Average	25% of results better, 50% of results worse
Dependability	0.68	Bad	In the range of the 25% worst results
Stimulation	0.81	Below Average	50% of results better, 25% of results worse
Novelty	0.66	Above Average	25% of results better, 50% of results worse

Figure 9.17: Table of UEQ Benchmark of old interface

Benchmarking the old interface against the the other products rated using the UEQ survey, it can be seen in Figures 9.15 and 9.17 that the old interface was rated Above Average for Perspicuity, Efficiency, and Novelty, Below Average for Attractiveness and Stimulation, and Bad for Dependability. Each of these scales were improved by at least one rank except for Attractiveness, which was improved by two ranks, and Novelty which remained the same. The increase in Attractiveness indicates that participants have an overall better impression of the New interface as compared to the Old interface.

9.5 Summary

There was a significant improvement in the system after the redesign in attractiveness and efficiency. Although the improvement in perspicuity, de-

pendability, and stimulation was found to be not statistically significant, there was an obvious trend towards significance. No significant difference was also found for Novelty between the old and new system. Future redesign phases could think of ways to incorporate new technology and ideas into the interface to increase the Novelty score for the system.

The next chapter will discuss my thoughts and opinions when applying the User-Centred Design methodology, as well as compare the old and new interfaces of LaBB-CAT and ElicitSpeech.

Chapter X

Conclusion

In this chapter, a summary of the thesis project will be outlined in Section 10.1. Section 10.2 will briefly describe my thoughts processes throughout the project whilst using a User-Centred redesign methodology. Finally, the current and future outlook for LaBB-CAT and ElicitSpeech will be detailed in Sections 10.3.1 and Section 10.3.2 respectively. Future visions are summarized in Section 10.4.

10.1 Summary

In this thesis project, a user-centred design methodology was applied for the redesign of a speech research database and a speech elicitation application. The design process utilized semi-structured interviews for the initial phases, and was changed to a questionnaire in the later phases. As the thesis project did not cover the entire redesign project, future design phases would still be conducted, culminating in the final implementation and evaluation of the designs.

Overall, the redesign project can be seen as successful. There was a marked improvement in the user interface for both LaBB-CAT and ElicitSpeech after the redesign, as indicated by the qualitative and quantitative results obtained respectively. The user interviews allowed the business clients to understand their users on a deeper level and find out the main problems in the interface designs.

Moreover, the redesign was well-received by both the project owners as well. For LaBB-CAT, one of the project owners, Robert Fromont, mentioned that the new search methodology is “really interesting” and really liked the separation of the search into Basic and Advanced searches. He also liked the

'helper' feature that allows for inserting common regular expressions, but it should be available for all non numeric layers and not only for phonemes. A few important things that are missing are the ability to anchor the words to the edges of turns or utterances, exporting audio, and using results in Praat TextGrids. The first appeared to be missing as none of the users interviewed had mentioned it as an important feature, and was therefore left out of the new interface. Exporting audio and Praat TextGrids would be implemented as a button to give users the option to use them. He also "really like" the button for exporting results to CSV without having to display the results first. Overall, he felt that the interface was "pretty cool", which indicates a successful project. The other project owner, Megan McAuliffe, stated that "As a relatively novice user of LaBB-CAT myself, I like the modifications made through the user-centered design process. Most importantly, it makes the search and export processes more efficient and easier to use."

For ElicitSpeech, Robert Fromont (one of the project owners) found the new interface to be "great" and "dynamic", while the other project owner, Megan McAuliffe liked the incorporation of encouragements and feedback, and found the interface to be "cool", "brighter", and "more visually appealing". She also liked the way that the consent form is integrated into a smaller text box, and the revised instructions delivery as they are "clear and more visually appealing". However, some things that she felt should be changed were wordings for the self assessment portion of speech fluency and motor disability, as well as the inclusion of birth years for older participants in their 70s to 80s.

10.2 User-centred design methodology

Through the application of the user-centred design methodology, there were many important learning points which could possibly help readers avoid the same mistakes as I did. One such point was regarding the allocation of time towards the different phases. I had previously allocated an equal amount of time toward each phase. However, on reflection, the amount of time taken towards each phase was different. The requirements gathering phase required the most time to complete, followed by the low fidelity phases, and then the

high fidelity phase. Better allocation of time would allow more gap time and better understanding of the time line for the project. Moreover, the amount of time taken for ethics application was longer than anticipated.

Another point learnt is that there are a huge number of software available on the market for prototyping from low-fidelity prototypes all the way to high-fidelity prototypes. Designers should explore the tools available to them before deciding which prototyping software to use for their project. Each software has unique pros and cons, and only by exploring their available features and functions can a designer decide what best suits their project.

I also learnt the importance of the sketches through my thesis project. Sketches allowed the flow of ideas, and are quick and easy to draw and visualise the design. The more sketches are created, the more ideas can be visualised and refined, and the better the final prototype will be.

10.3 Outlook

10.3.1 LaBB-CAT



Figure 10.1: Home page (Search) of new LaBB-CAT prototype

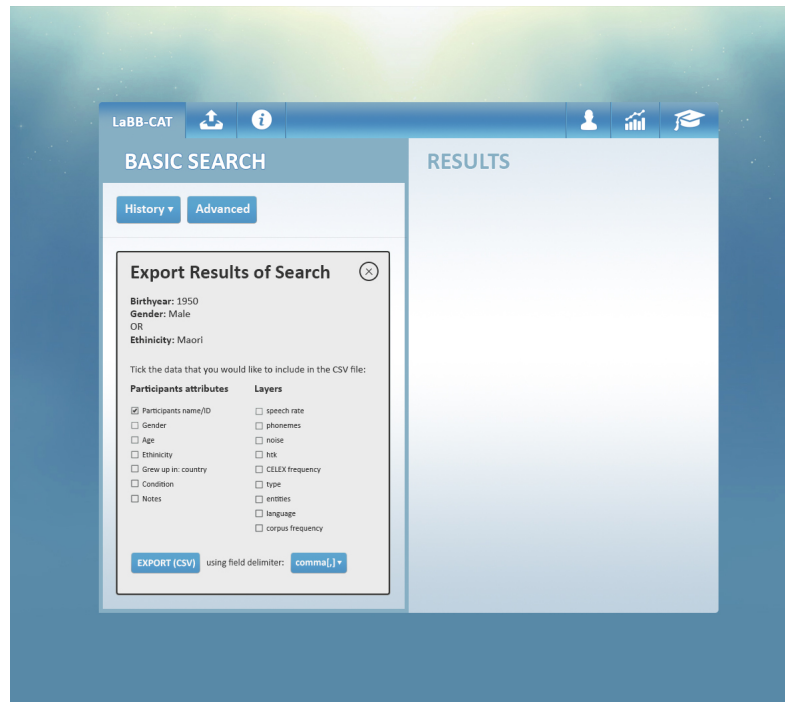
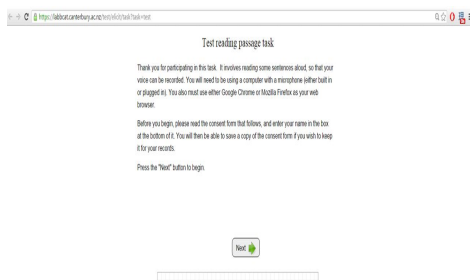


Figure 10.2: Home page (Search) of new LaBB-CAT prototype

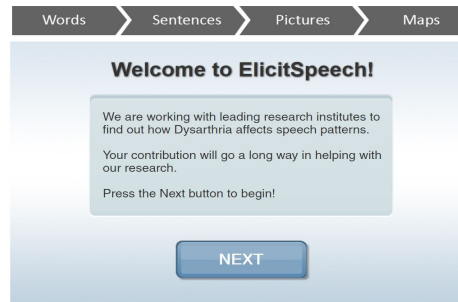
Comparing the changes in LaBB-CAT in the old and new interfaces (Figure 10.4), it can be seen that there has been a huge improvement in the interface after the redesign. Visual elements were added to the previous LaBB-CAT wireframe prototypes as shown in Figures 10.1, 10.2, and 10.3. As these changes have not been evaluated for its usability, future work for the redesign of LaBB-CAT would require a usability testing.

One of the main changes in future redesign phases would be to improve the learnability of LaBB-CAT for naive users. As there was a significant difference found in usability ratings between naive and experienced users, this indicates that the system needs to perhaps improve the intuitiveness of the interface, or improve the help system available for naive users. This would be a main area of focus going forward for LaBB-CAT.

Future uses for LaBB-CAT could also extend to all speech and linguistics researchers. One way could be to incorporate data analytics into the system, such as those used in Systematic Analysis of Language Transcripts (SALT) by speech researchers. As many participants interviewed had also mentioned the importance of SALT in their research and the overlapping functions that

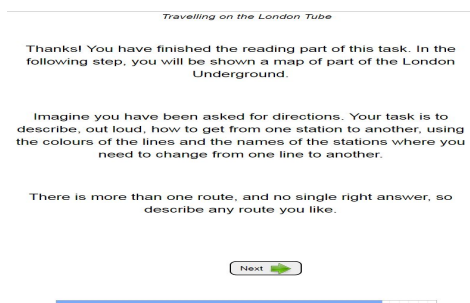


(a) Old interface of ElicitSpeech

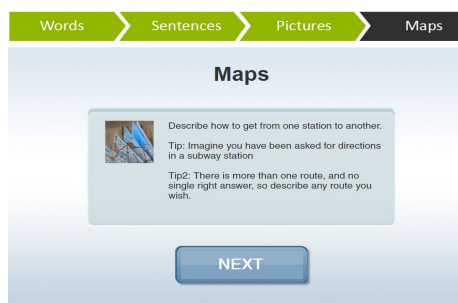


(b) New interface of ElicitSpeech

Figure 10.5: Home page of old interface (left) and new interface (right) of ElicitSpeech

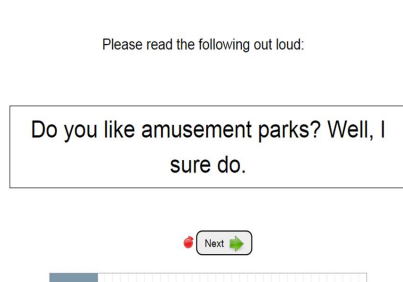


(a) Old interface of ElicitSpeech

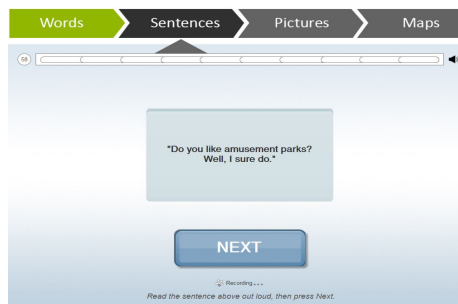


(b) New interface of ElicitSpeech

Figure 10.6: Instructions page of old (left) and new interface (right) of ElicitSpeech



(a) Old interface of ElicitSpeech

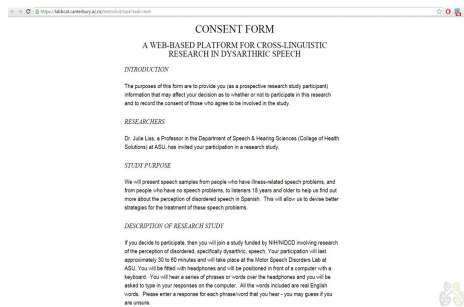


(b) New interface of ElicitSpeech

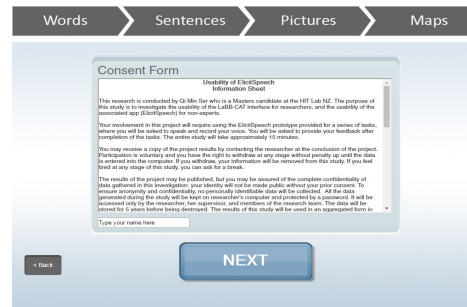
Figure 10.7: Task page of old (left) and new interface (right) of ElicitSpeech

10.3.2 ElicitSpeech

Figures 10.6, 10.7, and 10.8 shows a side-by-side comparison of the old interface with the new interface. Other than these direct comparisons, new pages were also added, such as the Motivation pages which encourage users after



(a) Old interface of ElicitSpeech



(b) New interface of ElicitSpeech

Figure 10.8: Consent page of old (left) and new interface (right) of ElicitSpeech

completion of a few tasks.

One of the main changes in future redesign phases would be to target the main target audience who would be using this application: Dysarthric patients, and the main promoters of this application: Speech Language Therapists. Although the main users of the application would be people with dysarthria, they would hardly use this application unless prompted by their clinical therapists. Therefore, if the application is sufficiently usable and user-friendly, the best forward strategy would be to promote the application to Speech Language Therapists and entice them to promote it to their patients instead.

Future uses for ElicitSpeech could also be to collect as much speech data from as many different people as possible. As the application is simply one that elicits speech from users for recording, it could be used as a tool that can be used for speech researchers to collect participants' speech data for future speech research. By increasing the amount of gamification incorporated into ElicitSpeech, it could be possible for users to promote the application to friends and increase traffic flow. ElicitSpeech could be akin to similar crowdsourcing citizen science platforms such as Zooniverse [22] and Galaxy Zoo [31], which spread rapidly through its users interest in contributing to research, as well as the gamification elements added to its tasks to make it appear less like menial work and more like fun games.

10.4 Visions for the future

Readers of this thesis will be able to learn about the process of a user-centred design approach, the benefits and shortcomings of using this approach, as well as the results that are produced. Other than allowing designers to create interfaces that specifically caters to the target audience, this thesis will also allow readers to understand how evaluation of an interface could be run using both low- and high-fidelity prototypes. The two applications redesigned in this thesis shows the difference between what users need and the interface that would be produced. As is the difference between day and night, this thesis highlights the difference in design, prototyping as well as evaluation solutions that are tailored for these two different applications.

My thesis project was successfully completed as the new interfaces designed were able to meet the requirements specified for both systems. This thesis project was able to create a usable interface of a speech research database, as well as an user-friendly interface of a speech recording application. The ease in using the speech research database might encourage more speech and linguistics researchers to use LaBB-CAT, as well as reduce the time taken for these users to find what they are looking for in the system. The improvement in attractiveness and efficiency of ElicitSpeech might improve users' willingness to use the application, which would increase the amount of data that could be collected using the application. These improvements would shorten the time taken for data collection, hence hastening the progress of future speech and linguistics research.

Future redesign for LaBB-CAT should focus on the learnability of the interface as well as the upload feature of the interface, while future redesigns for ElicitSpeech should focus on designing for the motoric disabilities that people with dysarthria might have, as well as to incorporate gamification elements to improve user motivation.

References

- [1] About UX | User Experience Professionals Association.
- [2] Definitions of User Experience and Usability | User Experience Professionals Association.
- [3] Don Norman - The Design of Everyday Things.
- [4] Measuring Usability with the System Usability Scale (SUS): MeasuringU.
- [5] Normans 7 Principles - HCI-06129.
- [6] NZILBB - LaBB-CAT.
- [7] System Usability Scale (SUS) | Usability.gov.
- [8] Treatment Efficacy Summaries.
- [9] What is Information Architecture? | IA Institute.
- [10] Statistics and Voice, Speech, and Language [NIDCD Health Information], June 2010.
- [11] 2012-2016 Strategic Plan Executive Summary | NIDCD, May 2012.
- [12] The Fundamental Checklist For Website Navigation Design & Architecture Part 1, July 2014.
- [13] Alan Amory, Kevin Naicker, Jacky Vincent, and Claudia Adams. The use of computer games as an educational tool: identification of appropriate game types and game elements. *British Journal of Educational Technology*, 30(4):311–321, October 1999.

- [14] Christoph Bartneck, Andreas Duenser, Elena Moltchanova, and Karolina Zawieska. Comparing the Similarity of Responses Received from Studies in Amazons Mechanical Turk to Studies Conducted Online and with Direct Recruitment. *PLOS ONE*, 10(4):e0121595, April 2015.
- [15] Sebastian Deterding. Gamification: designing for motivation. *interactions*, 19(4):14, July 2012.
- [16] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, pages 9–15. ACM, 2011.
- [17] C. B. Ferster and B. F. Skinner. *Schedules of reinforcement*, volume vii. Appleton-Century-Crofts, East Norwalk, CT, US, 1957.
- [18] David R. Flatla, Carl Gutwin, Lennart E. Nacke, Scott Bateman, and Regan L. Mandryk. Calibration games: making calibration tasks enjoyable by adding motivating game elements. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 403–412. ACM, 2011.
- [19] Robert Fromont and Jennifer Hay. ONZE Miner: the development of a browser-based research tool. *Corpora*, 3(2):173–193, November 2008.
- [20] Robert Fromont and Jennifer Hay. LaBB-CAT: An annotation store. In *Australasian Language Technology Association Workshop 2012*, volume 113. Citeseer, 2012.
- [21] Jesse James Garrett. *The Elements of User Experience*, March 2000.
- [22] Shannon Hall. Zooniverse Reaches One Million Volunteers, February 2014.

- [23] J. Hamari, J. Koivisto, and H. Sarsa. Does Gamification Work? A Literature Review of Empirical Studies on Gamification. In *2014 47th Hawaii International Conference on System Sciences*, pages 3025–3034, January 2014.
- [24] Zoe Kosmadoudi, Theodore Lim, James Ritchie, Sandy Louchart, Ying Liu, and Raymond Sung. Engineering design using game-enhanced CAD: The potential to augment the user experience with game elements. *Computer-Aided Design*, 45(3):777–795, March 2013.
- [25] Bettina Laugwitz, Theo Held, and Martin Schrepp. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and Usability Engineering Group*, pages 63–76. Springer, 2008.
- [26] Sus Lundgren and Staffan Bjork. Game mechanics: Describing computer-augmented games in terms of interaction. In *Proceedings of TIDSE*, volume 3, 2003.
- [27] Peter Morville. User Experience Design, June 2004.
- [28] Peter Morville and Louis Rosenfeld. *Information architecture for the World Wide Web*. O'Reilly, Sebastopol, CA, 3rd ed edition, 2007.
- [29] Scott Nicholson. A user-centered theoretical framework for meaningful gamification. *Games+ Learning+ Society*, 8(1):223–230, 2012.
- [30] Donald A. Norman. *The design of everyday things*. Basic Books, New York, New York, revised and expanded edition edition, 2013.
- [31] Jennifer Pinkowski. How to Classify a Million Galaxies in Three Weeks. *Time*.
- [32] Pier J.M. Prins, Sebastiaan DAVIS, Albert Ponsioen, Esther ten Brink, and Saskia van der Oord. Does Computerized Working Memory Training with Game Elements Enhance Motivation and Training Efficacy in Children with ADHD? *Cyberpsychology, Behavior, and Social Networking*, 14(3):115–122, March 2011.

- [33] Louis Rosenfeld, Peter Morville, and Jorge Arango. *Information Architecture: For the Web and Beyond, 4th Edition*. O'Reilly, Sebastopol, CA, 4th edition edition, 2015.
- [34] Dan Russell. Designer Information Search / Use Field Study Script, May 2015.
- [35] Erick Schonfeld. SCVNGR's Secret Game Mechanics Playdeck.
- [36] Martin Schrepp. User Experience Questionnaire Handbook, September 2015.
- [37] Gabe Zichermann and Christopher Cunningham. *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps*. "O'Reilly Media, Inc.", July 2011.

Appendix A

Appendix A: Interview script used in user study for LaBB-CAT

1. Background (5 mins)

- What kind of computer do you usually use?
- What kind of work environment do you have?
- What is your age?
- Tell us a little bit about what you do.
- Have you ever used LaBB-CAT or other speech research databases?
- How much time / week do you use the research database? (In hours.)
- Where do you usually use the Speech Research Databases?
- Do you use research databases on your mobile phone?(If so, what do you usually use it for?) (If so, remember to ask about this later on in the interview.)
- On a scale of 1 to 9, how much do you agree with this statement?
(1 Strongly Disagree, 9 Strongly Agree)
 - (a) I am very experienced in LaBB-CAT
 - (b) I am very experienced in Speech Research Databases

2. Research database background (8 min)

- Can you remember a few tasks you have had to do on the speech research databases? How often do you do that?

- What kinds of task do you usually do? What kinds of information are you looking for? For you? For others you work with?
 - How well does that task work on the research database? Do you usually have success?
 - What databases do you use for your speech research? Which database do you use the most? When did you first start using it? How did you hear about it?
 - Do you read helps, guides or forums to find how to use the database?
 - In your view, what is the most useful source of information for your work?
3. Questions for each task (5 minutes per task) We gave you a few sample tasks you would normally do on LaBB-CAT. When you're doing these tasks, use whatever websites/tools you'd like to use, and please think aloud.
- How do you decide what functions to use? (do they use the navigation, what words do they click on, what keywords/images do they look for ... etc...)
 - How much time would you want to spend on a task like this?
 - Are you satisfied with the outcome of the task? How do you decide if you have completed what you are trying to do?
 - How well do these methods work for you?
 - Would you have any difficulty re-doing that task again?
 - What was the main problem you had with this task?? (Say, "I need to find a particular kind of tool to use for my linguistics research" or If only there were some software to help me with this problem)
4. Feedback (SUS) (10 minutes)
- Would you like to use this system frequently? Why?
 - Do you find the system unnecessarily complex? Why?

- Do you think the system was easy to use? Why?
- Do you think that you would need the support of a technical person to be able to use this system? Why?
- Do you find the various functions in this system were well integrated? Why?
- Do you think that there was too much inconsistency in this system? Why?
- Do you think that most people would learn to use this system very quickly? Why?
- Do you find the system very cumbersome to use? Why?
- Do you feel very confident using the system? Why?
- Do you need to learn a lot of things before you could get going with this system? Why?
- What did you like about this system? Why?
- What did you dislike about this system?
- What do you think can be improved in this system? Why?
- Which function do you use most in LaBB-CAT?


5. Different kinds of task (5 mins)

- Can you tell me about a speech research task that took you a long time?
- Any tasks that you find yourself repeatedly doing all the time?
- Are there any kinds of task that you do NOT attempt to satisfy using research databases? If so, what do you do in these cases?

6. If you could create a magical tool that would allow you to complete any kind of speech research task you (plausibly) want, what would such a tool be? How would it work? What areas would it cover?

Appendix B

Appendix B: Questionnaires used in user study for LaBB-CAT



UC
UNIVERSITY OF
CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Please choose the option that best represents what you feel **for this task**.

Please decide spontaneously. Don't think too long about your decision to make sure that you convey your original impression.

Sometimes you may not be completely sure about your agreement with a particular attribute or you may find that the attribute does not apply completely to the particular system. Nevertheless, please tick a circle in every line.

It is your personal opinion that counts. Please remember: there is no wrong or right answer!

	Strongly disagree				Strongly agree
	1	2	3	4	5
This task was easy to do	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The task did not work well on the prototype	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am satisfied with the outcome of the task	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I spent more time than I expected on this task	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
These methods did not work well for me	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would have no problems re-doing this task again	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

What was the main problem you had with this task?

Any other comments about this task?

Please choose the option that best represents what you feel about the **entire system** you used.

Please decide spontaneously. Don't think too long about your decision to make sure that you convey your original impression.

Sometimes you may not be completely sure about your agreement with a particular attribute or you may find that the attribute does not apply completely to the particular system. Nevertheless, please tick a circle in every line.

It is your personal opinion that counts. Please remember: there is no wrong or right answer!

	Strongly disagree			Strongly agree	
	1	2	3	4	5
1. I think that I would like to use this system frequently	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. I found the system unnecessarily complex	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. I thought the system was easy to use	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. I think that I would need the support of a technical person to be able to use this system	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. I found the various functions in this system were well integrated	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. I thought there was too much inconsistency in this system	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. I would imagine that most people would learn to use this system very quickly	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. I found the system very cumbersome to use	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. I felt very confident using the system	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. I needed to learn a lot of things before I could get going with this system	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Appendix C

Appendix C: Interview script in user study for ElicitSpeech

1. Background (5 mins)
 - (a) How many hours approximately did you use a tablet (eg. iPad) this week?
 - (b) How many hours approximately did you use a computer or laptop this week?
 - (c) Have you ever designed or built an application for a tablet or website?
 - (d) What is your age?
2. Task (10 minutes) Lets try out the ElicitSpeech application now! Let them try it out!
3. Feedback (10 minutes)(Ask for Elaboration)
 - (a) Think about the home page of the system, what did you like and dislike about it?
 - (b) Think about the consent form section, what did you like and dislike about it?
 - (c) Think about the section where they asked you to enable your mike, what did you like and dislike?
 - (d) Think about the tasks where they ask you to read out loud a sentence, what did you like and dislike?

- (e) Think about the tasks where they ask you to describe map directions, what did you like and dislike?
 - (f) Did you like the last page of the application?
4. In general:
- (a) Would you like to use this system frequently?
 - (b) Do you find the system unnecessarily complex?
 - (c) Do you think the system was easy to use?
 - (d) Do you think that you would need the support of a technical person to be able to use this system?
 - (e) Do you find the various functions in this system were well integrated?
 - (f) Do you think that there was too much inconsistency in this system?
 - (g) Do you think that most people would learn to use this system very quickly?
 - (h) Do you find the system very cumbersome to use?
 - (i) Do you feel very confident using the system?
 - (j) Do you need to learn a lot of things before you could get going with this system?
 - (k) What did you like about this system?
 - (l) What did you dislike about this system?
 - (m) What do you think can be improved in this system?
5. I've been asking you a lot of questions. Is there anything you want to ask me?

Appendix D

Appendix D: UEQ used in user study 3 and 4 for ElicitSpeech

Please make your evaluation now.

For the assessment of the product, please fill out the following questionnaire. The questionnaire consists of pairs of contrasting attributes that may apply to the product. The circles between the attributes represent gradations between the opposites. You can express your agreement with the attributes by ticking the circle that most closely reflects your impression.

Example:

attractive ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ unattractive

This response would mean that you rate the application as more attractive than unattractive.

Please decide spontaneously. Don't think too long about your decision to make sure that you convey your original impression.

Sometimes you may not be completely sure about your agreement with a particular attribute or you may find that the attribute does not apply completely to the particular product. Nevertheless, please tick a circle in every line.

It is your personal opinion that counts. Please remember: there is no wrong or right answer!

Please assess the product now by ticking one circle per line.

	1	2	3	4	5	6	7		
annoying	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	enjoyable	1
not understandable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	understandable	2
creative	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	dull	3
easy to learn	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	difficult to learn	4
valuable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	inferior	5
boring	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	exciting	6
not interesting	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	interesting	7
unpredictable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	predictable	8
fast	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	slow	9
inventive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	conventional	10
obstructive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	supportive	11
good	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	bad	12
complicated	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	easy	13
unlikable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	pleasing	14
usual	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	leading edge	15
unpleasant	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	pleasant	16
secure	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	not secure	17
motivating	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	demotivating	18
meets expectations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	does not meet expectations	19
inefficient	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	efficient	20
clear	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	confusing	21
impractical	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	practical	22
organized	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	cluttered	23
attractive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	unattractive	24
friendly	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	unfriendly	25
conservative	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	innovative	26

Appendix E

Appendix E: Information sheet and consent forms

HIT Lab NZ
Telephone: +64 364 2349
Email: qi.ser@pg.canterbury.ac.nz
18.02.2016



Usability of LaBB-CAT Information Sheet

This research is conducted by Qi Min Ser who is a Masters candidate at the HIT Lab NZ. The purpose of this study is to investigate the usability of the LaBB-CAT interface for researchers, and the usability of the associated app (ElicitSpeech) for non-experts.

Your involvement in this project will require using the prototype for 5 tasks, which are tasks normally performed by researchers on research databases, and giving your feedback through an interview. You will be asked to think out loud whilst doing the tasks. The entire study will take approximately 60 minutes.

You may receive a copy of the project results by contacting the researcher at the conclusion of the project. Participation is voluntary and you have the right to withdraw at any stage without penalty up until the data is entered into the computer. If you withdraw, your information will be removed from this study. If you feel tired at any stage of this study, you can ask for a break.

The results of the project may be published, but you may be assured of the complete confidentiality of data gathered in this investigation: your identity will not be made public without your prior consent. To ensure anonymity and confidentiality, no personally identifiable data will be collected. All the data generated during the study will be kept on researcher's computer and protected by a password. It will be accessed only by the researcher, her supervisor, and members of the research team. The data will be stored for 5 years before being destroyed. The results of this study will be used in an aggregated form in Masters thesis of Qi Min Ser. A thesis is a public document and will be available through the UC Library.

The project is being carried out as a requirement for Masters degree by Qi Min Ser under the supervision of Christoph Bartneck, who can be contacted at christoph.bartneck@canterbury.ac.nz. He will be pleased to discuss any concerns you may have about participation in the project.

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee, and participants should address any complaints to The Chair, Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz).

If you agree to participate in the study, you are asked to complete the consent form and return it to the experimenter.

Qi Min Ser

HIT Lab NZ
Telephone: +64 3 64 2349
Email: qi.ser@pg.canterbury.ac.nz



Usability of LaBB-CAT Consent Form

I have been given a full explanation of this project and have had the opportunity to ask questions.

I understand what is required of me if I agree to take part in the research.

I understand that participation is voluntary and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided should this remain practically achievable.

I understand that any information or opinions I provide will be kept confidential to the researcher Qi Min Ser and that any published or reported results will not identify the participants. I understand that a thesis is a public document and will be available through the UC Library.

I understand that all data collected for the study will be kept in locked and secure facilities and/or in password protected electronic form and will be destroyed after ten years.

I understand the risks associated with taking part and how they will be managed.

I understand that I am able to receive a report on the findings of the study by contacting the researcher at the conclusion of the project.

I understand that I can contact the researcher Qi Min Ser (qi.ser@pg.canterbury.ac.nz) or supervisor Christoph Bartneck (christoph.bartneck@canterbury.ac.nz) for further information.

If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz)

By signing below, I agree to participate in this research project.

Name:

Date:

Signature: _____

Appendix F

Appendix F: Statistical tables

Tests of Normality							
InterfaceNfbcf		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Attractiveness	CF New	.162	19	.200 [*]	.907	19	.066
	CF Old	.181	14	.200 [*]	.967	14	.838
	FB New	.177	13	.200 [*]	.902	13	.140
	FB Old	.144	12	.200 [*]	.927	12	.352
Perspicuity	CF New	.181	19	.102	.897	19	.043
	CF Old	.126	14	.200 [*]	.965	14	.803
	FB New	.205	13	.138	.902	13	.143
	FB Old	.177	12	.200 [*]	.925	12	.334
Efficiency	CF New	.118	19	.200 [*]	.949	19	.383
	CF Old	.179	14	.200 [*]	.940	14	.421
	FB New	.179	13	.200 [*]	.901	13	.138
	FB Old	.163	12	.200 [*]	.929	12	.369
Dependability	CF New	.116	19	.200 [*]	.969	19	.754
	CF Old	.217	14	.073	.897	14	.100
	FB New	.116	13	.200 [*]	.928	13	.324
	FB Old	.292	12	.006	.752	12	.003
Stimulation	CF New	.228	19	.011	.906	19	.063
	CF Old	.161	14	.200 [*]	.954	14	.617
	FB New	.117	13	.200 [*]	.957	13	.703
	FB Old	.257	12	.028	.844	12	.031
Novelty	CF New	.189	19	.072	.933	19	.194
	CF Old	.132	14	.200 [*]	.977	14	.953
	FB New	.192	13	.200 [*]	.805	13	.008
	FB Old	.185	12	.200 [*]	.955	12	.707

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Figure F.1: Test of Normality for CrowdFlower and FaceBook data

Group Statistics					
InterfaceNfbcf		N	Mean	Std. Deviation	Std. Error Mean
Attractiveness	CF New	19	1.7721	1.09014	.25009
	FB New	13	1.3715	1.28988	.35775
Perspicuity	CF New	19	1.6711	1.06083	.24337
	FB New	13	1.5962	1.25639	.34846
Efficiency	CF New	19	1.5132	1.02223	.23452
	FB New	13	1.5962	1.12054	.31078
Dependability	CF New	19	1.1711	1.03095	.23652
	FB New	13	1.0000	1.27475	.35355
Stimulation	CF New	19	1.5526	1.17136	.26873
	FB New	13	1.0000	1.45057	.40232
Novelty	CF New	19	.9868	1.34751	.30914
	FB New	13	.2500	1.14564	.31774

Figure F.2: Descriptive statistics for CrowdFlower and FaceBook data of New Interface

CF vs FB New Interface										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
Attractiveness	Equal variances assumed	.919	.346	.948	30	.351	.40057	.42261	-.46252	1.26365
	Equal variances not assumed			.918	22.942	.368	.40057	.43650	-.50253	1.30366
Perspicuity	Equal variances assumed	.655	.425	.182	30	.857	.07490	.41144	-.76537	.91516
	Equal variances not assumed			.176	22.926	.862	.07490	.42504	-.80451	.95431
Efficiency	Equal variances assumed	.572	.455	-.217	30	.830	-.06300	.38249	-.86414	.69814
	Equal variances not assumed			-.213	24.304	.833	-.06300	.38934	-.88602	.72002
Dependability	Equal variances assumed	.826	.371	.419	30	.678	.17105	.40845	-.66311	1.00522
	Equal variances not assumed			.402	22.182	.691	.17105	.42537	-.71069	1.05280
Stimulation	Equal variances assumed	.597	.446	1.190	30	.243	.55263	.46443	-.39587	1.50113
	Equal variances not assumed			1.142	22.156	.266	.55263	.48381	-.45032	1.55559
Novelty	Equal variances assumed	2.923	.098	1.611	30	.118	.73684	.45734	-.19718	1.67086
	Equal variances not assumed			1.662	28.466	.107	.73684	.44332	-.17058	1.64427

Figure F.3: T-Test statistics for CrowdFlower and FaceBook data of New Interface

Group Statistics					
InterfaceNfbcf		N	Mean	Std. Deviation	Std. Error Mean
Attractiveness	CF Old	14	1.2736	.96804	.25872
	FB Old	12	.7775	.70828	.20446
Perspicuity	CF Old	14	1.1250	1.12980	.30195
	FB Old	12	1.2083	.99905	.28840
Efficiency	CF Old	14	.9643	.99931	.26708
	FB Old	12	.8542	.63477	.18324
Dependability	CF Old	14	.8929	.75774	.20252
	FB Old	12	.4375	.58509	.16890
Stimulation	CF Old	14	1.0179	1.03061	.27544
	FB Old	12	.5625	.56533	.16320
Novelty	CF Old	14	.9643	.84840	.22674
	FB Old	12	.3125	.81968	.23662

Figure F.4: Descriptive statistics for CrowdFlower and FaceBook data of Old Interface

CF vs FB Old Interface										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Attractiveness	Equal variances assumed	.923	.346	1.468	24	.155	.49607	.33785	-.20121	1.19335
	Equal variances not assumed			1.504	23.484	.146	.49607	.32976	-.18531	1.17745
Perspicuity	Equal variances assumed	.260	.615	-.198	24	.845	-.08333	.42167	-.95361	.78694
	Equal variances not assumed			-.200	23.966	.843	-.08333	.41755	-.94518	.77852
Efficiency	Equal variances assumed	2.231	.148	.329	24	.745	.11012	.33511	-.58150	.80174
	Equal variances not assumed			.340	22.284	.737	.11012	.32390	-.56110	.78134
Dependability	Equal variances assumed	1.764	.197	1.692	24	.104	.45536	.26910	-.10004	1.01075
	Equal variances not assumed			1.727	23.778	.097	.45536	.26370	-.08917	.99988
Stimulation	Equal variances assumed	2.512	.126	1.362	24	.186	.45536	.33423	-.23446	1.14518
	Equal variances not assumed			1.422	20.713	.170	.45536	.32016	-.21101	1.12173
Novelty	Equal variances assumed	.027	.870	1.983	24	.059	.65179	.32863	-.02647	1.33004
	Equal variances not assumed			1.989	23.623	.058	.65179	.32772	-.02517	1.32875

Figure F.5: T-Test statistics for CrowdFlower and FaceBook data of Old Interface

Tests of Normality							
Interface\I		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Attractiveness	Old	.136	26	.200 [*]	.981	26	.886
	New	.137	32	.133	.913	32	.013
Perspicuity	Old	.115	26	.200 [*]	.974	26	.736
	New	.143	32	.092	.914	32	.014
Efficiency	Old	.151	26	.131	.960	26	.388
	New	.136	32	.137	.934	32	.050
Dependability	Old	.191	26	.016	.933	26	.091
	New	.101	32	.200 [*]	.964	32	.362
Stimulation	Old	.171	26	.048	.952	26	.258
	New	.166	32	.025	.932	32	.046
Novelty	Old	.120	26	.200 [*]	.981	26	.892
	New	.111	32	.200 [*]	.965	32	.366

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Figure F.6: Test of Normality for Old and New interface

Group Statistics					
Interface\I		N	Mean	Std. Deviation	Std. Error Mean
Attractiveness	Old	26	1.0446	.87842	.17227
	New	32	1.6094	1.17219	.20722
Perspicuity	Old	26	1.1635	1.05105	.20613
	New	32	1.6406	1.12511	.19889
Efficiency	Old	26	.9135	.83649	.16405
	New	32	1.5469	1.04619	.18494
Dependability	Old	26	.6827	.70908	.13906
	New	32	1.1016	1.11958	.19792
Stimulation	Old	26	.8077	.86402	.16945
	New	32	1.3281	1.29894	.22962
Novelty	Old	26	.6635	.88301	.17317
	New	32	.6875	1.30291	.23032

Figure F.7: Descriptive statistics for Old Interface and New Interface

Independent Samples Test										
		Levene's Test for Equality of Variances		t-Test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Attractiveness	Equal variances assumed	5.299	.025	-2.035	56	.047	-.56476	.27756	-1.12078	-.00874
	Equal variances not assumed			-2.096	55.679	.041	-.56476	.26947	-1.10465	-.02487
Perspicuity	Equal variances assumed	.664	.419	-1.654	56	.104	-.47716	.28850	-1.05509	.10077
	Equal variances not assumed			-1.666	54.867	.101	-.47716	.28644	-1.05123	.09691
Efficiency	Equal variances assumed	2.857	.097	-2.504	56	.015	-.63341	.25301	-1.14025	-.12658
	Equal variances not assumed			-2.562	55.992	.013	-.63341	.24722	-1.12865	-.13818
Dependability	Equal variances assumed	5.747	.020	-1.655	56	.103	-.41887	.25302	-.92573	.08799
	Equal variances not assumed			-1.732	53.112	.089	-.41887	.24189	-.90401	.06627
Stimulation	Equal variances assumed	8.450	.005	-1.751	56	.085	-.52043	.29723	-1.11595	.07499
	Equal variances not assumed			-1.824	54.072	.074	-.52043	.28538	-1.09256	.05169
Novelty	Equal variances assumed	3.341	.073	-.080	56	.936	-.02404	.29963	-.62426	.57618
	Equal variances not assumed			-.083	54.399	.934	-.02404	.28816	-.60167	.55360

Figure F.8: T-Test statistics for Old Interface and New Interface